



El tractament arxivístic de Twitter

Aniol Maria Vallès

Director del treball: Joan Soler Jiménez

Any d'elaboració: 2014

Treball de recerca del Màster d'Arxivística i Gestió de Documents

Escola Superior d'Arxivística i Gestió de Documents

Col·lecció Treballs fi de màster i de postgrau

Com citar aquest article: Maria Vallès, Aniol. (2014) *El tractament arxivístic de Twitter*. Treball de recerca del Màster d'Arxivística i Gestió de Documents de l'Escola Superior d'Arxivística i Gestió de Documents. (Treballs fi de Màster i de postgrau). [Http://...](http://...) (consultat el ...)



Aquesta obra està subjecta a llicència Creative Commons Reconeixement-NoComercial-SenseObraDerivada 3.0 Espanya (<http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.ca>). Es permet la reproducció total o parcial i la comunicació pública de l'obra, sempre que no sigui amb finalitats comercials, i sempre que es reconegui l'autoria de l'obra original. No es permet la creació d'obres derivades.

Resum

Aquest treball tracta sobre el tractament de la documentació electrònica generada a Twitter, és a dir, l'accessibilitat i gestió de la documentació generada en aquesta xarxa social. Quines opcions ens ofereix Twitter i quin tractament podem fer del que accedim a Twitter.

Al llarg dels diversos apartats s'exposa i analitza tant la informació a través de les metadades que ens ofereix Twitter com també quines propostes s'han plantejat des de diversos àmbits i institucions per a treballar sobre aquesta font d'informació i dades. Fins arribar a determinar què entenem per documentació de Twitter i en quina forma l'hem de mentalitzar o amb quines eines podem ser capaços de provar de treballar-hi.

Paraules clau: documentació electrònica, xarxes socials, microblogs, arxius digitals, metadades, formats web, JSON, XML, estructures de dades, llenguatges extensibles.

Título: El tratamiento archivístico de Twitter

Resumen

Este trabajo se propone explicar el tratamiento de la documentación electrónica generada en Twitter, la accesibilidad y gestión de la documentación generada en esta red social. Qué opciones nos ofrece Twitter y qué tratamiento podemos hacer de lo que accedemos en Twitter.

A lo largo de los diversos apartados se expone y analiza la información a través de los meta-datos que nos ofrece Twitter como también qué propuestas se han planteado des de distintos ámbitos e instituciones para trabajar sobre esta fuente de información y datos. Para llegar a determinar qué entendemos por documentación de Twitter y en qué forma la podemos mentalizar o con qué herramientas podemos ser capaces de probar a trabajar con ella.

Palabras clave: documentación electrónica, redes sociales, microblogs, archivos digitales, meta-datos, formatos web, JSON, XML, estructuras de datos, lenguajes extensibles.

Title: The archivist treatment of Twitter

Abstract

This essay is about the treatment of electronic documents generated in Twitter, the accessibility and management of the documentation generated in this social website. What options offers Twitter and the treatment that we can do on what we access at Twitter.

Along the different sections we will expose and analyze the information through the meta-data that offers Twitter and also the proposals that have been brought in the different areas and institutions for working on this source of information and data. And at the end, we will try to determine what we understand as documentation of Twitter and in which form we have to mentalising it or with which tools we can be able to try to work on it.

Keywords: electronic documents, social websites, microblogs, digital archives, meta-data, web formats, JSON, XML, markup languages, data structures, extensible languages.

SUMARI

1	Introducció.....	7
2	Marc teòric i estat de la qüestió.....	10
2.1	Twitter i la API.....	10
2.1.1	Què és Twitter.....	10
2.1.2	Què és una API.....	11
2.2	Dificultats del treball.....	13
2.3	Diplomàtica amb Twitter.....	14
2.3.1	Anàlisi diplomàtica de Twitter.....	15
	Context.....	15
	Sistema Jurídic.....	16
	Funció.....	18
	Acte o fet.....	19
	Formalització.....	20
	Forma.....	21
	Essència.....	25
	Propietats Essencials.....	25
	Explicar més coses amb les metadades JSON.....	27
2.4	L'estudi de Twitter.....	32
2.4.1	La credibilitat de Twitter.....	32
2.4.2	Implicació a Twitter.....	35
2.4.3	Els diaris personals i Twitter.....	39
2.4.4	Històries amb Twitter.....	40
3	Cos de la recerca: Casos arxivístics amb Twitter.....	41
3.1	L'exemple de NARA.....	41
3.2	L'exemple dels National Archives del Regne Unit.....	47
3.3	L'exemple de la Facultat de Biblioteconomia i Documentació.....	48
3.4	Quina podria ser la meua proposta.....	50
3.4.1	El model OAIS i iArxiu.....	56
4	Conclusions.....	66
4.1	Objectius i preguntes.....	66
4.2	Recorregut del treball.....	67
4.3	Projectes a dur a terme.....	68
4.4	Agraïments.....	68
5	Glossari.....	69
6	Bibliografia i fonts.....	74
6.1	Documentació de Twitter.....	74
6.2	NARA i Library of Congress.....	75
6.3	Altres articles.....	76
6.4	Llibres.....	78
7	Annexos.....	79
7.1	Manual per extreure dades en JSON amb Postman i l'API de Twitter.....	79
7.1.1	Les metadades d'un tuit.....	86
7.1.2	Les metadades d'un perfil.....	88
7.2	Descarregar el nostre arxiu d'usuari de Twitter.....	90

1 Introducció

En el món dels documents segurament veuríem com cadascú té una concepció del document delimitada, amb els seus marges, el seu contingut, la seva diferenciació. Estem previnguts però que en la nova era digital en que ja ens anem sentint immersos, tota aquesta concepció s'ha de veure qüestionada. Estem davant de la des-materialització del document. L'accés a la informació i no als documents. En el món informàtic es perd la idea del suport. Per això en el món anglòfon es parla de mitjans en comptes de suports. I per això des de ja fa anys trobem reflexions com la de Montserrat Sebastià i Salat que ja ens deien que la «transmissió d'informació comporta comunicació, i la comunicació comporta actualment des-materialització o virtualització de la documentació i de la informació»¹.

La meua inquietud per l'anàlisi de la documentació electrònica i les meta-dades que l'acompanyen s'inicia cap a finals de 2012 quan a l'assignatura de Diplomàtica del Màster en Arxivística i Gestió de Documents a l'ESAGED apreng l'anàlisi de documents a partir del Context, la Forma i l'Essència. Això em desenvolupa la lògica d'aplicar aquest anàlisi més enllà dels documents electrònics més comuns com podrien ser les fotografies digitals o els documents en PDF, per pensar que encara es podria fer un pas més i buscar la seva aplicació en la documentació generada dels missatges de les xarxes socials.

En un principi els meus coneixements no miraven més enllà de l'anàlisi en cru del codi HTML que mostra un navegador qualsevol. No serà fins que exposi la meua idea en una pàgina de micro-blogging que compartia amb un grup reduït², que apreng que a més a més del PHP i SQL hi ha una eina que es diu API i que serveix per trametre-li ordres per a que et retorni informació. En aquells inicis amb la API podies amb tan sols una URL i diverses variables que hi afegies, obtenir directament en la finestra del navegador el conjunt de dades en format JSON que formaven el document anomenat tuit, més tard com ja explicarem la tècnica es va complicar i les meta-dades s'han anat ampliant també.

A l'any 2012 Twitter anuncià quasi al Nadal³, que activava una nova funció als usuaris per a que poguessin demanar i descarregar-se un "arxiu" (#TwitterArchive) amb tots els seus tuits publicats fins llavors. El que et facilitaven però era la descàrrega d'un fitxer en format ZIP, en el qual hi contenien 3 tipus de presentacions. Per una banda un format accessible i fàcil per a la majoria

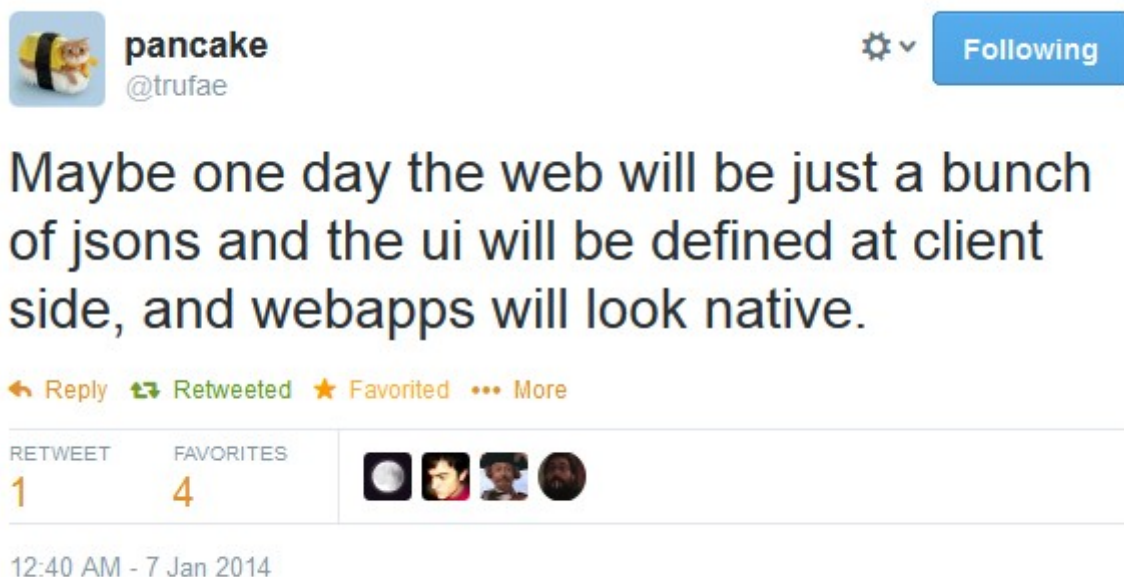
1 SEBASTIÀ, M., (1998). "Archiveros/infotecnólogos: la Internet como motor del cambio de imagen de los archiveros". A: El archivo en el entorno cultural. Coslada: Comunidad de Madrid, p. 91-112.
<http://anabad.org/images/publicaciones/anabad/xiiijarchivosmunicipales.pdf> (consultat a 23 d'abril de 2014)

2 En concret s'anomenava Jisko (<http://jisko.org>) i va ser considerada una de les primeres xarxes de micro-blogging en castellà. Vaig entrar en el projecte per fer-ne de traductor al català i vaig acabar col·laborant amb d'altres tasques com la gestió de l'entorn Wiki per exemple.

3 Vandor, Mollie. "Your Twitter archive | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2012/your-twitter-archive> (consultat a May 1, 2014)

d'usuaris que era un fitxer en HTML que servia per visualitzar en el nostre navegador una interfície que recordava a Twitter i que permetia cercar i visualitzar en el propi navegador web tuits concrets i demés. Per l'altra s'oferia també dues opcions de visualització de les dades dels tuits, per JSON i per CSV, opcions que permetien aquestes sí visualitzar les meta-dades que acompanyaven els tuits, i que tenien emmagatzemats mes per mes i any per any tots els tuits en format .js i .csv respectivament.⁴ Avui dia sembla que encara continua oferint-se aquest servei i va millorant-se. Ja que per exemple ara el format en CSV ja s'envia dins un sol arxiu .csv (abans era en varis fitxers distribuïts per anys i mesos), amb tots els tuits introduïts curiosament en orde invers, és a dir el primer en aparèixer és l'últim que vam publicar a Twitter.

Segurament si expliquéssim tot això a un informàtic ens miraria amb incredulitat i es preguntaria en quina realitat ens hem aïllat tots aquests anys, ja que tot aquest procés ja porta quasi des dels inicis de la informàtica. És més, no fa gaires mesos m'arribava a través de Twitter aquest plantejament⁵:



1. Traducció: *Potser algun dia la web serà només un manat de jsons i la interfície serà definida només per la banda del client, i les aplicacions web lluiran natives.*

Segurament a vista ràpida costaria d'entendre per al públic profà el plantejament exposat, però si ho expliquéssim d'una manera més planera, el missatge del tuit es refereix a que tot i que la gent quan consulta Internet veu imatges, text, vídeos, i escolta música, per sota de tot això, en les entranyes hi corre un codi anomenat web ("teranyina") que entrellaça els elements per mantenir-los en un conjunt al que anomenen pàgines web. Aquest codi en els inicis d'Internet

4 Als Annexos a "Descarregar el nostre arxiu d'usuari de Twitter" hi hem posat un senzill manual de com es pot obtenir aquest "arxiu" del nostre compte de Twitter i quines possibilitats ofereix.

5 Twitter / trufae: Maybe one day the web will ... <https://twitter.com/trufae/status/420339088831442944> (consultat a 22 de març de 2014)

s'estandarditzaria en el llenguatge HTML (4.01 seria la versió més longeva), més tard apareixerien les webs dinàmiques i moltes webs es farien amb el llenguatge PHP (que al final seria visualitzat en HTML en els navegadors), per tornar en l'actualitat a un codi més estàtic però alhora extensible anomenat HTML5. Doncs amb l'aparició d'encara més llenguatges de programació que facilitarien l'elaboració de noves webs anomenades ara xarxes socials, veuríem també la difusió de nous llenguatges de marcatge per tal de transmetre la informació en forma d'objectes. Un d'aquests llenguatges és el JSON⁶ que anomena el missatge del tuit, i per tant el que ens està opinant aquest tuit és que potser en un futur no massa llunyà tot Internet ja correrà amb llenguatges de marcatge que cada màquina interpretarà i permetrà visualitzar de manera nativa.

Podríem evocar a *Urkundendämmerung* o mite de la mort documental de que sovint parla Joan Soler⁷. La representació en els dispositius multimèdia no ens acaba de servir, no ens mostra elements per demostrar la fiabilitat o autenticitat del document, el document surt mutilat? No ben bé, surt adaptat a les circumstàncies. Ja no està intrínsecament associat a la seva representació, sinó a la procedència certa de les dades (el que en diríem la base de dades de Twitter). El document són parts que es mostren per un canal o varien en un altre segons les peticions i intencions de difusió, les porcions de contingut que s'estructuren en argument i es comuniquen com a informació. Hem de demandar si volem les dades i metadades completes ja que no tindrem de per sí una forma estable. «Sólo los datos serán enviados, la procedencia y el origen de estos deberá ser comprobada, en esta comprobación residirá la credibilidad de los nuevos sistemas»⁸.

Però no hi donem tantes voltes i determinem ja quina base té aquest treball. Hem de determinar què és el document del que volem parlar en aquest treball, i es tracta dels tuits. Tuits que estan en una base de dades relacionada, i que podem rebre'n porcions en format JSON. Aquest fragment de la base de dades en JSON i que anomenem "tuit"⁹ seria un document simple. En canvi el conjunt de tuits que formen la base de dades a Twitter seria el document compost.

Volem tractar el tuit com a document, però no en la seva visió més popular, del format amb que surt als navegadors, sinó en la seva forma més interna, en les seves entranyes, allà on podem trobar la informació bàsica del document i no només la seva forma. Amb una simple imatge d'una

6 Un altre podria ser el XML que molts arxivers ja coneixen. Segurament si hem fet edició de textos coneixerem el LaTeX, i si hem treballat en edició web el Markdown.

7 Soler Jiménez, Joan. "Diplomàtica.cat." *Urkundendämmerung o el mito de la muerte documental*. <http://diplomaticapuntcat.blogspot.com.es/2013/02/urkundendammerung-o-el-mito-de-la.html> (consultat a 6 de maig de 2014)

8 Ídem

9 Segons el Termcat el vocabulari indicat per a anomenar l'objecte tant podria ser: «piulada, n f tuit, n m sin. compl.

Missatge curt, amb un nombre de caràcters limitat, publicat instantàniament en un microblog.

Nota: El sinònim complementari tuit prové de Twitter, que és el nom comercial de l'aplicació que gestiona la plataforma de microblogs més utilitzada actualment.»

captura de pantalla veuríem però no sabríem moltes de les dades que ens responen a perquè el tuit pot ser tractat com a document en arxivística. Volem aprofundir en el coneixement tant de les dades que ens arriben en el navegador com de les metadades que acompanyen al tuit i que són accessibles des de la API. Com diu Michael Buckland, es tracta no només de rebre el document sinó d'entendre'l, tenir-lo en un llenguatge adequat, en un nivell adequat i ser coherent¹⁰.

2 Marc teòric i estat de la qüestió

2.1 Twitter i la API

2.1.1 Què és Twitter

Twitter és una de les primeres xarxes socials en desenvolupar el que s'anomenaria microblog. El seu naixement es deu a Jack Dorsey, Evan Williams, Biz Stone i a Noah Glass, que tot i crear el sistema el març del 2006 no van publicar la primera versió del lloc web fins al juliol de 2006. I des de llavors el lloc web va anar guanyant ràpidament popularitat a tot el món, amb 500 milions d'usuaris el 2012, a un ritme actual de 340 milions de tuits per dia¹¹.

La idea del sistema de microblog va sorgir de quan els creadors de Twitter treballaven a Odeo, una companyia de distribució de podcasts, i necessitaven un sistema de comunicacions internes. És així com d'un sistema de missatges breus interns es va adoptar el microblog. El microblog es tractava de minimitzar l'estructura d'un blog per tal de fer-lo el més directe i ràpid de llegir possible, per tal que la informació arribés el més ràpid i clara possible. Per això el que es va fer és prendre com a model el que llavors era el missatge breu més estès (fora de les trucades perdudes que tan sols indicaven tocs d'atenció¹²), i es tractava dels SMS. És a dir, un missatge breu de 140 caràcters. Això va afavorir que les comunicacions fossin molt ràpides i que la gent al estar molt acostumada als SMS entrés fàcilment en la idea del microblog de Twitter. Fins a convertir-se actualment en sinònim d'immediatesa i informació actual.

En el nostre panorama cultural però, malgrat que fou ràpidament i de manera molt transversal

10 Buckland, Michael (2013). «La qualitat de la informació al web». BiD: textos universitaris de biblioteconomia i documentació, desembre, núm. 31. <http://bid.ub.edu/31/buckland1.htm> (consulta 10 de maig de 2014)

11 Wikipedia contributors, "Twitter," Wikipedia, The Free Encyclopedia, <http://en.wikipedia.org/w/index.php?title=Twitter&oldid=607331920> (consultat a 6 de maig de 2014)

12 Per si alguna persona jove no sap què són, el Termcat les descriu com:

« trucada perduda, n f
es llamada perdida
en lost call
en missed call

<Terminologia>

Trucada que es talla abans de poder establir comunicació.»

difós el seu ús entre els catalanoparlants¹³, i tot i les reticències¹⁴, Twitter no publicà la versió de la web en català (traduïda a més per 5000 voluntaris¹⁵) fins al 6 de maig de 2012¹⁶. Així que tot i que des del 2007 ja apareixien personatges rellevants de la cultura i el món internauta publicant tuits a Twitter, no és fins al 2012 que hi ha l'entrada més massiva i definitiva de la cultura catalana a Twitter per exemple.

El que genera Twitter són bàsicament dues coses: Perfils i Tuits. És a través d'aquests dos elements que es construeix la xarxa social de microblog. Els perfils són els comptes registrats dels usuaris que tenen la voluntat d'anar publicant periòdicament tuits, i els tuits són els missatges curts de 140 caràcters que ja hem anomenat tant a la Introducció com més amunt en aquest apartat. Podem extreure dades i metadades dels dos elements, tot i que Twitter ens ofereix més variables per extreure més informació com per exemple de cerques, però aquesta informació que s'extreu és la que està fixada en els dos elements que hem anomenat, els tuits i els perfils. Nosaltres ens fixarem més en els tuits, ja que són el que més atrau de Twitter, el que es genera en més quantitat i el que aporta encara més varietat d'informació. Són els tuits a més els que més canvis han vist per part de Twitter ja que el que seria els Perfils dels usuaris han mantingut bàsicament les mateixes dades exposades des de quasi els inicis de Twitter¹⁷.

Com ja hem dit a la introducció d'aquest treball ens fixarem en el que forma el tuit, no pas en la seva aparença externa que apareix en els navegadors web, aplicacions i d'altres canals de difusió de Twitter. Ja que entenem que el document essencial es troba en les dades de fons i les metadades que les acompanyen, les quals tècnicament es troben a les bases de dades de Twitter i a la que se'ns permet l'accés tutelat a través de la API.

2.1.2 Què és una API

Una API és un sistema d'interacció entre l'usuari d'un servei web i els sistemes del servei web, o com ho defineix Termcat un «conjunt de rutines que utilitza una aplicació per a demanar i executar determinats serveis realitzats per un sistema operatiu»¹⁸. Un sistema que ens permet dotar-nos de

13 Araguás, Natalia. "El català brilla a Twitter." Ara.cat. http://www.ara.cat/xarxes/Twitter-catala-llengues-mapa_0_592140899.html (consultat a 6 de maig de 2014)

14 Ventura, Patricia. "Albert Cuesta: "Twitter no quiere traducir al catalán y debe de tener algún motivo"." LA VANGUARDIA. <http://www.lavanguardia.com/internet/20110922/54219937926/albert-cuesta-twitter-no-quiere-traducir-al-catalan-y-debe-de-tener-algun-motivo.html> (consultat a 6 de maig de 2014)

15 Araguás, Natalia. "El català brilla a Twitter." Ara.cat. http://www.ara.cat/xarxes/Twitter-catala-llengues-mapa_0_592140899.html (consultat a 6 de maig de 2014)

16 I. Gómez, Laura. "Twitter now in Catalan and Ukrainian | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2012/twitter-now-in-catalan-and-ukrainian> (consultat a 6 de maig de 2014).

17 Amb la versió actual els perfils mostren la data de registre dels usuaris, el nombre de favorits, entre d'altra informació que potser al 2006-2007 no es mostrava tan fàcilment, però diguem-ne que de 2006 a 2013 la variació de la informació dels perfils no ha variat quasi.

18 «API

la capacitat de fer ús d'un servei web que no és nostre, dins d'una aplicació i de manera segura. Al implementar-la l'API et posa a disposició certs mètodes o funcions per al teu desenvolupament d'aplicacions i programes. Tot el desenvolupament que es realitzi estarà limitat pels mètodes o funcions que aquesta API inclogui, és a dir que no podrem afegir noves funcionalitats. És així com Twitter pot decidir i assegurar-se de què fem i com ho fem.¹⁹

Aquesta interacció es fa a través d'una línia d'ordres o comandes. Proveeix una interfície de programació d'aplicacions (*Application Programming Interface*), un conjunt d'indicacions sobre les funcions i procediments a través de biblioteques o programari. És el terreny d'interacció entre qui farà l'ús informàtic i els serveis de què farà ús. Una API és una part d'un Kit de Desenvolupament d'una Aplicació (*SDK*)²⁰, i en general ofereix varis mitjans per al desenvolupament entre varis llenguatges de programació.

Per aquest treball hem hagut d'aprendre com treballar amb la nova versió de l'API de Twitter, la qual com exposem més endavant ha variat al llarg del temps i a cada versió requereix d'una major implicació i interacció per part de l'usuari aportant codis identificatius i de verificació. Alhora també van sorgint noves maneres de treballar-hi amb més o menys usabilitat adaptant-se a les demandes dels usuaris, sobretot les de codi obert.

En el cas de Twitter l'accés a la seva API es fa a través de la web "Twitter Application Management"²¹ a la que ens hem de registrar a través de OAuth (el sistema de registre d'usuaris propi de Twitter basat en el sistema OpenID). Un cop registrats haurem de crear una aplicació per tal d'obtenir les claus d'accés al sistema de consultes i peticions de la API. La documentació que ens indicarà quines ordres podrem executar es troba a la web "Twitter Developers"²² i proveeix de llistats per a cada tipus d'ordre i petició amb tota la informació per a dur-les a terme²³.

es interfaz de programación de aplicaciones
 es API sigla
 fr interface de programmation
 fr interface de programmation d'applications
 fr interface de programme d'application
 fr API sigla
 en application program interface
 en application programming interface
 en API sigla

Conjunt de rutines que utilitza una aplicació per a demanar i executar determinats serveis realitzats per un sistema operatiu.

Nota: La sigla API correspon a l'equivalent anglès application programming interface ('interfície de programació d'aplicacions').»

19 Rodríguez, Andrea. "Qué es una API." Bitelia. <http://bitelia.com/2014/05/que-es-api> (consultat a 23 de maig de 2014)

20 Col·laboradors de Viquipèdia, "Interfície de programació d'aplicacions," Viquipèdia, l'Enciclopèdia Lliure, [//ca.wikipedia.org/w/index.php?title=Interf%C3%Adcie_de_programaci%C3%B3_d%27aplicacions&oldid=13247266](http://ca.wikipedia.org/w/index.php?title=Interf%C3%Adcie_de_programaci%C3%B3_d%27aplicacions&oldid=13247266) (descarregat 20 d'abril de 2014)

21 "Twitter Application Management." Twitter. <https://apps.twitter.com> (consultat a 13 de maig de 2014)

22 "Twitter Developers." Twitter. <https://dev.twitter.com> (consultat a 13 de maig de 2014)

23 Als Annexos d'aquest treball hi hem posat un tutorial "Manual per extreure dades en JSON amb

2.2 Dificultats del treball

Les dificultats d'aquest treball però són que no disposem de totes les dades ni ens és legalment possible fer-ho. Només la Biblioteca del Congrés dels Estats Units d'Amèrica disposarà d'aquestes dades quan es faci efectiu la cessió dels tuits que Twitter va firmar amb ells el 14 d'abril de 2010 i que des del 4 de gener de 2013 ja se n'ha sabut bona part²⁴, però poc més.

Twitter ha continuat donant mostres d'estar treballant en oferir més facilitats d'accés a les dades per part dels usuaris quan el 15 d'abril de 2014 va anunciar que havia comprat Gnip²⁵, un sistema que recull dades estadístiques i socials del que es publica a Twitter, amb xifres com 2,3 trilions de tuits analitzats de 42 països als que ofereix la informació que en recopila.

Per altra banda, trobem l'enorme obsessió per generar estadístiques amb Twitter. Segurament una via de futur molt útil i atraient per a publicistes, però que per no desviar-nos del nostre objectiu del treball, hem procurat no centrar-nos-hi massa. A més, com demostra el contracte de cessió de dades de Twitter i la Llibreria del Congrés²⁶, a la col·lecció que s'arxivarà s'hi emmagatzemarà la informació dels tuits i es deixa el tema de la recopilació d'estadístiques i dades socials als investigadors futurs.

Nosaltres hem volgut centrar-nos en el tema de la documentació generada al publicar a Twitter, no en la documentació generada a partir del que s'ha publicat a Twitter. Com ja hem dit, ens centrarem en la forma documental i els elements que conformen aquesta documentació relacionada amb el que anomenem tuits. El que després es pugui generar a partir de l'agrupació d'aquests documents i el seu tractament estadístic ens portaria cap a un altre terreny en el que no volem divagar en aquest Treball Final de Màster.

Si volem doncs definir ben bé de què estem parlant haurem d'ajudar-nos de dos conjunts teòrics com són l'Arxivística i la Diplomàtica. Primer per la Diplomàtica ja que és la que ens permet analitzar a partir de la unitat més petita i poder establir amb precisió els elements que conformen el document. Per després passar a quin tractament a nivell de conjunt arxivístic se n'ha fet llavors valorant les propostes d'Arxivística. Ja que com que no disposem de masses dades de com es disposarà la cessió de Twitter a la Llibreria del Congrés, excepte que es determinarà en forma de col·lecció, no podem parlar massa sobre quins aspectes tècnics conformaran la forma arxivística en que acabaran a l'abast del públic els tuits de Twitter.

Postman i l'API de Twitter" sobre com interactuar i enviar ordres o peticions a aquesta API de Twitter.

24 "Update on the Twitter Archive at the Library of Congress." Library of Congress Blog. <http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/> (consulta 16 d'abril de 2014).

25 "Twitter welcomes Gnip to the flock | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2014/twitter-welcomes-gnip-to-the-flock> (consultat a 16 d'abril de 2014)

26 "Library of Congress - Twitter" GIFT AGREEMENT <http://blogs.loc.gov/loc/files/2010/04/LOC-Twitter.pdf> (consultat a 18 de març de 2014)

2.3 Diplomàtica amb Twitter

Començant per la definició de Diplomàtica, ens diu el DIEC²⁷ que «la diplomàtica és la disciplina que té per objecte l'estudi de la forma, del procés d'elaboració i de la tradició dels diplomes antics». Per tant es queda només amb documents antics.

En canvi José Ramón Cruz Mundet parla de com «la diplomàtica és el tractat del document que l'analitza críticament en tot el seu conjunt, però essencialment en la seva forma, gènesis, evolució, tradició i conservació, disposant aquest anàlisi a l'ordre d'esbrinar la seva autenticitat, saber si és veritable o fals, si ho és en tot o en part, i establir el seu valor com a font històrica»²⁸.

I si consultem la Viquipèdia²⁹ la definició també parla de documentació de manera més “àmplia”, i diu que «la diplomàtica és la ciència dels documents o diplomes [...] s'interessa per l'estudi d'aquells documents amb una natura i un valor jurídic, administratiu, històric, o testimonial, així com pel coneixement de diverses formalitats que li confereixen un caràcter probatori i una força legal». Quedem llavors que no cal que ens quedem en la documentació antiga, i podem agafar la documentació recent. Així que si podem agafar documentació recent, podríem agafar-ne també de la que es produeix electrònicament a les xarxes socials, per exemple la de Twitter.

Segurament trobarem reticències i qui dirà que encara és aviat per valorar la rellevància de Twitter, però el que percebem és que des del president d'Estats Units d'Amèrica (@BarackObama) fins al de Veneçuela (@chavezcandanga³⁰) han utilitzat profusament aquesta xarxa social com a mitjà de comunicació ràpida i directa amb els seus ciutadans i seguidors, i fins i tot el Papa de Roma (@Pontifex) ha acabat volent aprofitar-se d'aquesta xarxa social.

També sabem que dins dels Estats Units d'Amèrica s'està donant molta rellevància a l'activat d'aquesta xarxa social, fins al punt que la Biblioteca del Congrés dels Estats Units va comprar l'arxiu de Twitter, amb tots els tuits des del 2006³¹. Compra que no sabem massa quin sentit tindrà si qualsevol usuari ja pot descarregar-se la seva part de tuits publicats a Twitter³². A més a més, també podríem anomenar com l'Administració Nacional d'Arxius i Enregistraments (NARA) “ha llançat algunes tecnologies i signat acords per crear Arxius Nacionals de llocs com Twitter,

27 Institut d'Estudis Catalans - Diec2 <http://dlc.iec.cat/results.asp?txtEntrada=diplom%E0tica&OperEntrada=0> (consultat a 16 d'abril de 2014)

28 CRUZ MUNDET, José Ramón. *Archivística. Gestión de documentos y administración de archivos*. Alianza Editorial, 2012. pg. 105

29 Col·laboradors de Viquipèdia, "Diplomàtica," Viquipèdia, l'Enciclopèdia Lliure, <http://ca.wikipedia.org/w/index.php?title=Diplom%C3%A0tica&oldid=11289698> (descarregat 11 de març de 2013)

30 Després de la seva mort es va decidir no publicar-hi res més, per ara. L'últim tuit anunciava en to personal un 18 febrer de 2013 estar en mans del personal mèdic de l'hospital.

31 La Biblioteca del Congreso de Estados Unidos adquiere el archivo de Twitter | Tecnología | EL PAÍS http://tecnologia.elpais.com/tecnologia/2010/04/14/actualidad/1271235670_850215.html (consultat a 16 d'abril de 2014)

32 Twitter has started rolling out the option to download all your tweets - The Next Web <http://thenextweb.com/twitter/2012/12/16/twitter-has-started-rolling-out-the-option-to-download-all-your-tweets/> (consultat a 16 d'abril de 2014)

YouTube, Facebook, i Flickr, i és NARA qui està usant aquestes eines i llocs per crear noves maneres de comunicació amb el públic³³.

En el nostre cas d'estudi amb la Diplomàtica parlarem del cas de Barack Obama. Perquè concretament d'ell? Doncs perquè és un personatge polític que sempre acaba tenint més ressò en les xarxes que cap altre dels seus opositors polítics³⁴, i fins i tot encapçala el llistat del més referenciat a la pròpia xarxa de Twitter³⁵ malgrat no ser ell qui té més seguidors a Twitter³⁶. També el triem a ell perquè diferencia els missatges del seu usuari de Twitter que són personals seus, dels que són més "impersonals", i ho fa amb una marca "—President Obama" (anteriorment "-bo") que ja analitzarem més endavant la seva rellevància, i els diversos aspectes que conformen un tuit amb aquesta marca que segons ell designa el que ell personalment diu. Per tant el nostre objecte d'estudi serà la diplomàtica d'un tuit d'Obama, si segueix alguna mena de protocol o què podem extreure d'informació per tal de saber sobre quina mena de document ens ofereix quan publica un tuit.

Així doncs ens disposarem a definir els elements i conceptes bàsics (de context, forma i essència) que permeten completar un anàlisi diplomàtic del document anomenat tuit.

2.3.1 Anàlisi diplomàtica de Twitter

Context

Hem d'entendre el context dins d'un període ampli o no, ja que Obama ja utilitzava noves tecnologies com la Blackberry des de feia anys, i n'era un gran fan³⁷. I de fet sembla que segueix pensant i tenint molt present la marca de la companyia RIM fins i tot avui dia³⁸. La limitada oferta d'aplicacions i atacs que estan orientats a les ja no tant populars Blackberry fa que no es qüestioni massa la seva elecció.³⁹

Però la seva eclosió a les xarxes socials la podríem determinar a la campanya popularment coneguda com "Yes We Can"⁴⁰, que ens portaria moltes pàgines d'explicació i que podríem

33 Social Media and Web 2.0 <http://www.archives.gov/social-media/> (consultat a 16 d'abril de 2014)

34 Los tuits de Barack Obama impactan más que los de Mitt Romney - Tecnología - CNNMéxico.com <http://mexico.cnn.com/tecnologia/2012/11/07/los-tuits-de-barack-obama-impactan-mas-que-los-de-mitt-romney> (consultat a 16 d'abril de 2014)

35 Golden Tweets <http://2012.twitter.com/en/golden-tweets.html> (consultat a 16 d'abril de 2014)

36 Katy Perry <https://twitter.com/katyperry> per exemple, va pels 52 milions de seguidors en aquests moments. Quasi 10 milions més que Obama.

37 La publicidad que le hace Obama a BlackBerry no tiene precio - lanacion.com <http://www.lanacion.com.ar/1090720-la-publicidad-que-le-hace-obama-a-blackberry-no-tiene-precio> (consultat a 16 d'abril de 2014)

38 Barack Obama Still Uses a BlackBerry - Techvibes.com <http://www.techvibes.com/blog/barack-obama-still-uses-a-blackberry-2012-09-11> (consultat a 16 d'abril de 2014)

39 Blitzer, Wolf. "No 'Candy Crush' for President Obama - CNN.com Video." CNN. <http://edition.cnn.com/video/?/video/tech/2014/05/23/wolf-intv-erica-fink-obama-blackberry.cnn-ap> (consultat a 28 de maig de 2014)

40 Wikipedia contributors, "Barack Obama presidential campaign, 2008," Wikipedia, The Free Encyclopedia, <http://en.wikipedia.org/w/index.php?>

resumir-ho amb que va fer una explotació al màxim de la interactivitat amb les xarxes socials i la figura del candidat a president. No seria massa agosarat dir que va ser l'eclosió simbòlica de la difusió de campanyes electorals a través de les xarxes socials. Mètode que molts han volgut imitar però que cap ha aconseguit un impacte similar, i menys a Europa, malgrat els esforços que a vegades s'han fet per anomenar la "Revolució Islandesa"⁴¹ com un altre projecte per l'estil, que com entre d'altres, no està tan clar.

Nosaltres concretament analitzem un context més posterior, i és el de la re-elecció d'Obama com a president dels EUA. I com novament va muntar una xarxa d'intercomunicacions aprofitant-se de tot tipus de xarxes socials, i fins i tot construint tot un entramat dedicat exclusivament a estudiar el comportament, l'operativitat i aprofitament d'aquestes xarxes socials per tal de fer arribar el missatge polític al màxim de votants possibles⁴². I assolint novament la presidència dels EUA. Qui sap si ajudat altra vegada per aquest impuls de les xarxes socials, però el fet és que hi va recórrer de nou, fins i tot endinsant-se a pàgines de discussió molt obertes com Reddit⁴³, demostrant novament el bon assessorament i coneixença de saber "on es mou" l'activisme a Internet, i arribant a fer caure el sistema de servidors d'aquesta web⁴⁴ degut a l'entusiasme dels usuaris.

És en tot aquest context on hi va jugant un paper destacat el Twitter d'Obama (recordar que a Facebook també ronda pels 40 milions de seguidors⁴⁵ no més). I és en el seu Twitter on hi va publicar els missatges breus, les coses més immediates: el que deia en els seus mítings⁴⁶, o a quina web estava responent als seus electors⁴⁷. I per això se'n cuidava prou de distingir-lo respecte d'altres llocs web on també hi era present. I era a Twitter on firmava com ja hem assenyalat els seus propis missatges amb la marca "-bo" per exemple, en cap xarxa social més (que sapiguem).

Sistema Jurídic

Bé, seria difícil parlar del sistema jurídic a Internet, malgrat que s'han fet molts intents de posar

title=Barack_Obama_presidential_campaign,_2008&oldid=603971638 (consultat a 16 d'abril de 2014)

41 Wikipedia contributors, "2009 Icelandic financial crisis protests," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=2009_Icelandic_financial_crisis_protests&oldid=603813560 (consultat a 16 d'abril de 2014).

42 El arma total de Obama para vencer a Romney: un megacerebro informático que controla Facebook - Público.es <http://www.publico.es/internacional/445113/el-arma-total-de-obama-para-vencer-a-romney-un-megacerebro-informatico-que-controla-facebook> (consultat a 16 d'abril de 2014)

43 Barack Obama surprises internet with Ask Me Anything session on Reddit | Technology | guardian.co.uk <http://www.guardian.co.uk/technology/us-news-blog/2012/aug/29/barack-obama-ask-me-anything-reddit> (consultat a 16 d'abril de 2014)

44 Obama responde preguntas en Reddit - FayerWayer <http://www.fayerwayer.com/2012/08/obama-responde-preguntas-en-reddit/> (consultat a 16 d'abril de 2014)

45 Barack Obama <https://www.facebook.com/barackobama/likes> (consultat a 16 d'abril de 2014)

46 Twitter / BarackObama: Watch live: President Obama's ... <https://twitter.com/BarackObama/status/265663193777569792> (consultat a 16 d'abril de 2014)

47 Twitter / BarackObama: "Hi, I'm Barack Obama, President ... <https://twitter.com/BarackObama/status/241209767983472641> (consultat a 16 d'abril de 2014)

normes⁴⁸. Però excepte casos amb molta ironia⁴⁹, la majoria cauen en poca operativitat ja que la xarxa d'Internet en principi és neutral⁵⁰. A més, els sistemes d'anonimat com TOR⁵¹ també funcionen per entrar a Twitter, o fins i tot sorgeixen sistemes per publicar tuits encriptats⁵² (que per contra han de ser breus però són possibles d'enviar i no ser fàcilment desxifrats). Cosa que fa bastant difícil la operativitat d'un sistema jurídic en els casos de més alt nivell de secretisme.

Però podríem parlar de quina normativa regeix Twitter per exemple, i quin sistema jurídic, judicial i executiu contempla. Per a entrar en aquest apartat però haurem d'endinsar-nos a la web de Twitter, en concret a l'apartat TOS o Termes del servei d'aquesta web⁵³:

En aquest apartat Twitter ens diu que l'ús de la seva web, serveis i publicació de continguts respondrà en que nosaltres hem acceptat els seus termes, ja que en teoria així ho hem indicat també al registrar-nos a la web. Per tant, "signem el contracte" de termes del servei només accedint-ne a l'ús de l'eina. Twitter però ens fa responsables de tot el que nosaltres publiquem a la seva web, i de les conseqüències que se'n puguin ocórrer. Si ets dins d'una companyia, organització, entitat o país t'has de responsabilitzar tu de complir també la normativa del lloc des d'on estiguis enviant el que publiques a Twitter.

A més, Twitter no es responsabilitza de la pèrdua de servei, ni tampoc tenen perquè avisar-te de cap interrupció que succeeixi, tant sigui del servei com de qualsevol altre característica de l'aplicació.

A més Twitter es reserva el dret a introduir-te publicitat dirigida a tu, o a proporcionar informació de tu als anunciants que ho reclamin, sobretot respecte als continguts que tu envies públicament a la web. En temes de privacitat Twitter es reserva el dret a retenir dades teves obeint a la seva "Pòlissa de Privacitat"⁵⁴, pòlissa que també especifica que Twitter compleix amb la U.S.-EU i els principis de la U.S.-Swiss Safe Harbor Privacy, en el tema de seguretat, integritat de les dades, accés i demés.

48 Wikipedia contributors, "Stop Online Piracy Act," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Stop_Online_Piracy_Act&oldid=603478166 (consultat a 16 d'abril de 2014)

49 Rules of the Internet | Know Your Meme <http://knowyourmeme.com/memes/rules-of-the-internet> (consultat a 16 d'abril de 2014)

50 Wikipedia contributors, "Net neutrality," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Net_neutrality&oldid=527843815 (consultat a 16 d'abril de 2014)

51 Wikipedia contributors, "Tor (anonymity network)," Wikipedia, The Free Encyclopedia, [http://en.wikipedia.org/w/index.php?title=Tor_\(anonymity_network\)&oldid=527522936](http://en.wikipedia.org/w/index.php?title=Tor_(anonymity_network)&oldid=527522936) (consultat a 16 d'abril de 2014)

52 Ja veurem el temps que duren donat que no són gaire extensament usats i desapareixen tan ràpid com n'apareixen de nous, però avui podem anomenar per exemple un dels més publicitats per Anonymous que és "FIGHT CENSORSHIP!! being more safe on social networking sites... AnonTwi <http://anontwi.sourceforge.net>" (consultat a 15 de juny de 2014)

53 Twitter / Twitter Terms of Service <https://twitter.com/tos> (consultat a 16 d'abril de 2014)

54 Twitter / Twitter Privacy Policy <https://twitter.com/privacy> (consultat a 16 d'abril de 2014)

Twitter no es fa responsable de la veracitat, o de la cura i fiabilitat del que els usuaris publiquen als seus perfils de la xarxa social. Per tant no es farà responsable si un mateix en fa mal ús i decideix publicar dades confidencials o transmetre alguna dada privada públicament.

A canvi de tot això rebràs el dret a usar, copiar, reproduir i usar on vulguis el que has publicat a Twitter. És a dir se't permetrà distribuir els teus tuits pel mitjà que et plaguin. A més també rebràs el dret que Twitter et proveirà, promourà, i millorarà el servei i la manera de publicar continguts, per tal de facilitar-te la tasca de difondre allò que t'interessi. Això es farà sobretot a través de la API⁵⁵.

Twitter però també es reserva el dret de ser ells mateixos qui usin, copiïn, i reproduïxin el que tu has publicat. Per tant t'has de responsabilitzar de què has publicat ja que no se't garanteix quin abast de distribució posterior podria tenir el que has fet públic. La llicència del que es publica a Twitter queda sota Copyright, com també ho està la major part de l'aplicació web, logotips i demés continguts oficials de Twitter.

Es penalitzarà però fer-se passar per un personatge real que un no és, és a dir la suplantació d'identitat⁵⁶. També es respectarà que les marques, noms i companyies comercials puguin reclamar sobre qui estigui fent servir la seva identitat sense permís oficial. També es penalitzarà la publicació de dades confidencials de persones, com adreces de carrers, o números d'identitat sense el permís o l'autorització de la persona identificada. També s'anirà contra les conductes violentes, el SPAM, i d'altres atacs directes contra el sistema que perjudiquin el servei i els usuaris.

Finalment Twitter avisa que es reserva el dret de desactivar qualsevol compte o servei que cregui oportú, i que també et dona dret a tu a desactivar el teu perfil si així ho prefereixes. El pacte dels termes d'ús és de tu a tu, per tant si acceptes dones a entendre que has entès els termes i que per tant se't poden aplicar les normes que s'han descrit en els termes.

En principi es respondrà sota la llei de l'Estat de Califòrnia, a més de la legislació local dels usuaris, però primerament Twitter recorrerà a la llei del Comtat de San Francisco i dels Estats Units d'Amèrica, ja que és on hi té la seu l'empresa. Per acabar se'ns diu que la última revisió d'aquests termes va entrar en fase efectiva el 25 de juny de 2012, per tant fa uns dos anys que no es revisen. I que si es vol, es pot consultar les variacions que se n'han anat fent fins arribar a l'actual⁵⁷.

Funció

Com ja hem dit la Funció del perfil a Twitter és la de recopilar el que un va fent en el moment més

55 Developer Rules of the Road | Twitter Developers <https://dev.twitter.com/terms/api-terms> (consultat a 16 d'abril de 2014)

56 Twitter Help Center | The Twitter Rules <https://support.twitter.com/articles/18311-the-twitter-rules> (consultat a 16 d'abril de 2014)

57 Twitter / Twitter Terms of Service <https://twitter.com/tos/previous> (consultat a 16 d'abril de 2014)

immediat possible, i la Funció d'un tuit per la seva banda és la de comunicar. Podríem determinar que els tuits són comunicacions i llavors els perfils esdevindrien registres de comunicacions.

Hem d'anar amb compte amb les comunicacions de Twitter. Hem d'evitar no creuar massa el marge entre la privadesa i la privacitat, ja que potser a un no li interessa gaire que el rastregin punt per punt sobre tot allò que fa. És per això que sovint els personatges públics es posen algun marge entre publicació i publicació a Twitter. Al llarg d'aquests últims anys havíem vist exemples molt curiosos com per exemple el del Papa Benet XVI que es va posar els dimecres com a dia per publicar el que ell creia oportú⁵⁸. Tot i que en canvi Obama (@BarackObama) però no té cap dia marcat, sinó que va fent dia a dia. Això sí amb unes hores marcades que ens permeten veure com està de controlat el sistema per tal que no estigui 24 hores enviant coses i no sembli un robot, sinó que es pugui veure en quines hores "el Twitter del president dorm"⁵⁹.

Acte o fet

L'acte o fet ja hem vist que té una certa legislació que és d'obligat compliment un cop s'entra a Twitter. I el ritual bé, diguem que el Twitter d'Obama vol mostrar una imatge esporàdica del president, en escenes públiques més o menys informals⁶⁰, d'altres més formals i protocol·làries⁶¹, sense desapropiar el moment de publicar també entremig algun missatge polític més directe⁶².

Òbviament té tot un equip que li fa les fotografies i tot. Sabem com cada fotografia oficial i els moments més importants d'Obama són supervisats abans de ser publicats a qualsevol xarxa social. Sobretot va ser destacable el cas de la fotografia d'Obama a Flickr⁶³ que mostrava a Obama i demés membres del govern assistint al relat de l'operació de captura de Bin Laden l'1 de maig de 2011.

Per tant veiem com el Fet, l'Acció i els Agents estan molt establerts. Fins i tot les composicions visuals, l'enfocament, o la situació dels personatges, res queda trivial a la Casa Blanca⁶⁴. En aparença, i si no que ho preguntin a l'expresident espanyol José Luis Rodríguez Zapatero quan va voler portar les seves filles a la Casa Blanca i aquestes van sortir exposades no

58 La agencia española que gestiona la cuenta del Papa en Twitter vive este miércoles como "un día para la historia" <http://www.europapress.es/sociedad/noticia-agencia-espanola-gestiona-cuenta-papa-twitter-vive-miercoles-dia-historia-20121212192011.html> (consultat a 16 d'abril de 2014)

59 Barack Obama @BarackObama's Sleeping Schedule <http://www.sleepingtime.org/BarackObama> (consultat a 16 d'abril de 2014)

60 Twitter / BarackObama: One of our favorite 2012 photos: ... <https://twitter.com/BarackObama/status/278932381224816640> (consultat a 16 d'abril de 2014)

61 Twitter / BarackObama: In step. <http://t.co/x2OSOxhP> <https://twitter.com/BarackObama/status/278650641205850112> (consultat a 16 d'abril de 2014)

62 Twitter / BarackObama: Taking care of family: Just ... <https://twitter.com/BarackObama/status/278581141605347328> (consultat a 16 d'abril de 2014)

63 Wikipedia contributors, "Situation Room," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Situation_Room&oldid=527831923 (consultat a 16 d'abril de 2014)

64 Los re medios: La foto más visitada en la historia de Flickr y porqué <http://los-re-medios.blogspot.com/2011/05/la-foto-mas-visitada-en-la-historia-de.html> (consultat a 16 d'abril de 2014)

de manera gaire protocol·lària al Flickr oficial de la Casa Blanca, i el rebombori que va portar el tema a Espanya⁶⁵.

Formalització

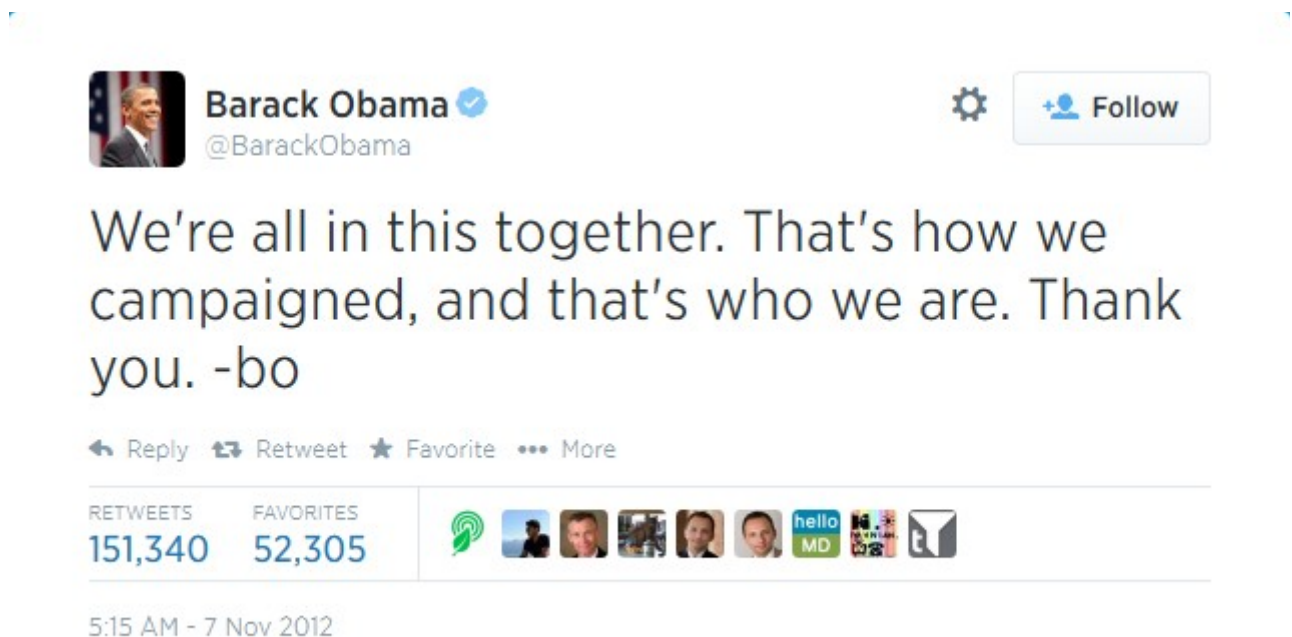
Podem mostrar la capçalera identificativa del perfil d'Obama a Twitter. Amb el seu avatar identificatiu, la marca de compte verificat, el seu sobrenom, la seva biografia, la localització que s'ha volgut posar, i una adreça web. En la part inferior el número de tuits (*TWEETS*) al moment que hem fet la captura, el número de perfils que segueix (*FOLLOWING*), i el número de perfils que el segueixen (*FOLLOWERS*).



1. Il·lustració: <https://twitter.com/BarackObama>

⁶⁵ La polémica fotografía de las hijas de Zapatero - RTVE.es
<http://www.rtve.es/noticias/20090925/polemica-fotografia-hijas-zapatero/293874.shtml> (consultat a 16 d'abril de 2014)

Però passem ja a analitzar el nostre objectiu. I per això hem triat el següent tuit:



2. Il·lustració: <https://twitter.com/BarackObama/status/266031109979131904>

Com podem veure és un tuit bastant personal. Primer de tot perquè es dirigeix el missatge a tota la gent que va acompanyar Obama en la seva campanya electoral, els agraeix la feina feta i els té en record. Després veiem que va firmat amb “-bo” (que ara firmaria com hem dit abans amb “— President Obama”), és a dir es tracta d'un tuit d'Obama personal, no és cap dels que podem veure en el seu perfil sobre fotografies “espontànies” en actes públics o sobre missatges polítics concrets.

Forma

Unitat documental composta o simple

Com a forma podríem determinar que hem triat una Unitat Documental Simple, ja que tan sols hem agafat un tuit concret, i no pas tot el perfil sencer de l'Obama que llavors potser ja tractaríem com a Unitat Documental Composta. Si recordem, establim que el tuit és una comunicació i el perfil un registre de comunicacions.

Ens hagués servit de molta ajuda saber com la Biblioteca del Congrés dels Estats Units està tractant el tema⁶⁶, però suposem que 167TB no deuen ser fàcils de classificar (ja avisaran diuen quan ho tingui enllestit⁶⁷). Així doncs nosaltres hem suposat aquesta divisió entre UDS i UDC, comunicació i registre de comunicacions. Però tenim constància que el procés s'està fent⁶⁸, i

66 Twitter as History - Library of Congress Signs Up - NYTimes.com <http://www.nytimes.com/2010/04/15/technology/15twitter.html> (consultat a 16 d'abril de 2014)

67 The Library and Twitter: An FAQ | Library of Congress Blog <http://blogs.loc.gov/loc/2010/04/the-library-and-twitter-an-faq/> (consultat a 16 d'abril de 2014)

68 Our Online Privacy Policy | The White House <http://www.whitehouse.gov/privacy> (consultat a 16 d'abril de 2014)

que passarà a formar part de l'arxiu del president o algun apartat per l'estil, malgrat que no s'especifiqui ben bé a on ho trobarem.

Procediment

El procediment de com s'ha constituït el tuit el desconeixem bastant. Podríem saber qui és el fotògraf si sortís alguna fotografia, o podríem saber quin protocol se segueix al publicar coses a Twitter si mostressin algun vídeo, a l'estil de com ho fa el Vaticà⁶⁹. Però res, hem intentat fins i tot contactar amb la Casa Blanca i no hi ha hagut resposta de cap mena. Així que ens tocarà especular amb els elements essencials més endavant.

Forma física

Com veiem la representació del tuit és en format digital, llenguatge HTML que sabem que prové d'una programació en llenguatge Ruby⁷⁰. I tota la informació s'emmagatzema en una base de dades MySQL⁷¹. Tenim també el Javascript accessible sota la marca de Bootstrap⁷² entre d'altres codis dins la web. Comptant també la fulla d'estils CSS que organitza i dona un disseny vistós al conjunt d'elements que es mostren al navegador.

La forma intel·lectual

En el tuit la forma intel·lectual és un text encapsulat, amb una part superior reservada pel nom i el sobrenom de qui envia el tuit junt amb la marca de *verified account*⁷³ (la icona blava i blanca), un botó per opcions varies sobre aquest autor (enviar-li un missatge privat *DM* per exemple) i un botó per seguir-lo.

Tot seguit veiem el text format que s'ha enviat (en anglès òbviament). Just a sota 3 enllaços per a respondre el tuit, retuitejar-lo, o marcar-lo com a favorit. Sota d'això el número de retuits, favorits, i un breu llistat de persones que o han marcat a favorit o ho han retuitejat.

Per últim ens apareix l'hora, la data i un enllaç per veure les opcions d'incrustar aquest tuit en d'altres espais web. La data i l'hora estan enllaçades amb la direcció URL del tuit, la qual ens oferirà la sèrie de números que constitueixen la identificació d'arxivament del tuit.

Aquesta forma intel·lectual actualment ha variat ja que s'han aplicat varies actualitzacions en el disseny de Twitter i els usuaris han anat rebent aquests canvis gradualment. Però bàsicament és bastant similar a el que expliquem en aquest apartat.

69 Pope tweets for the first time - YouTube <http://www.youtube.com/watch?v=gy4sNYmfpt8> (consultat a 16 d'abril de 2014)

70 ree « snax <http://blog.evanweaver.com/2009/09/24/ree/> (consultat a 16 d'abril de 2014)

71 How Twitter tweets your tweets with open source | ZDNet <http://www.zdnet.com/how-twitter-tweets-your-tweets-with-open-source-7000003526/>

72 Bootstrap <http://twitter.github.com/bootstrap/> (consultat a 16 d'abril de 2014)

73 "FAQs about verified accounts." Twitter Help Center. <https://support.twitter.com/groups/31-twitter-basics/topics/111-features/articles/119135-about-verified-accounts> (consultat a 16 d'abril de 2014)

Vincle diplomàtic

Podem veure com el format és el típic format de Twitter, i si es dubtés ja hem inclòs també a URL a la llegenda de la imatge per tal que sigui consultable a la pròpia web de Twitter i es pugui verificar que sí es tracta d'un tuit de Twitter. Veiem com surten els elements identificatius amb la fotografia d'Obama i el seu sobrenom, juntament amb la marca de *verified account*. Per tant ens va confirmant que prové d'on diu provenir.

Contingut

«We're all in this together. That's how we campaigned, and that's who we are. Thank you. -bo»⁷⁴. És un missatge "original", només va a aparèixer al compte de Twitter d'Obama.

Tampoc som uns experts en l'escriptura d'Obama, ni en els seus discursos, però el missatge podríem pensar que va bastant en sintonia amb el que sovint expressa públicament Obama als seus seguidors. No demostra un to estrident o diferenciat "del que ens té acostumats".

Tradició documental

Clar que ara podríem trobar còpies repartides en vàries webs, perquè estarien copiant el text que es va publicar en el seu dia a Twitter. Però podríem dir que és un missatge original, dissenyat només per aquesta xarxa, fet reforçat per la característica brevetat del comunicat que pretén ser breu i concís en el missatge, i no va anar acompanyat de cap comentari més en el seu moment. Per tant es tractava d'una píndola solta en la multitud de missatges, fotografies i tot el que vindria després amb la investidura d'Obama.

Perfil documental

Al ser un document electrònic podríem trobar vàries vies per mostrar les metadades del document que serien un reflex del seu perfil documental. En aquest cas hem triat JSON ja que ens ha semblat prou clara i és la que ofereix més fàcilment Twitter.

Així doncs el codi JSON del document (tuit) triat seria el següent⁷⁵:

```
{
  created_at: "Wed Nov 07 04:15:33 +0000 2012"
  id: 266031109979131900
  id_str: "266031109979131904"
  text: "We're all in this together. That's how we campaigned, and that's who we are. Thank you. -bo"
  source: "web"
  truncated: false
  in_reply_to_status_id: null
```

74 "We're all in this together. That's how we campaigned, and that's who we are. Thank you. -bo " <https://twitter.com/BarackObama/status/266031109979131904>

75 Anys abans amb una URL https://api.twitter.com/1/statuses/show.json?id=266031109979131904&include_entities=true en teníem prou per obtenir l'esquema en JSON de totes les metadades del tuit. Ara però el procés és més complex i requereix d'alguns coneixements informàtics més elevats com són els de funcions i peticions a una API, a través d'alguna mena de terminal o línia d'ordres. En l'apartat Annexos detallem en el manual "Manual per extreure dades en JSON amb Postman i l'API de Twitter" com es pot obtenir tot aquest esquema de metadades en JSON.

```

in_reply_to_status_id_str: null
in_reply_to_user_id: null
in_reply_to_user_id_str: null
in_reply_to_screen_name: null
-user: {
  id: 813286
  id_str: "813286"
  name: "Barack Obama"
  screen_name: "BarackObama"
  location: "Washington, DC"
  description: "This account is run by Organizing for Action staff. Tweets from
the President are signed -bo."
  url: http://t.co/O5Woad92z1
  -entities: {
    -url: {
    -urls: [
    -{
      url: http://t.co/O5Woad92z1
      expanded_url: http://www.barackobama.com
      display_url: "barackobama.com"
    -indices: [
      0
      22
    ]
    }
    ]
    }
  -description: {
    urls: [ ]
  }
}
protected: false
followers_count: 42538034
friends_count: 651575
listed_count: 206111
created_at: "Mon Mar 05 22:08:25 +0000 2007"
favourites_count: 4
utc_offset: -14400
time_zone: "Eastern Time (US & Canada)"
geo_enabled: false
verified: true
statuses_count: 11585
lang: "en"
contributors_enabled: false
is_translator: false
is_translation_enabled: true
profile_background_color: "77B0DC"
profile_background_image_url:
http://pbs.twimg.com/profile_background_images/451819093436268544/kLbRvwBg.png
profile_background_image_url_https:
https://pbs.twimg.com/profile_background_images/451819093436268544/kLbRvwBg.png
profile_background_tile: false
profile_image_url:
http://pbs.twimg.com/profile_images/451007105391022080/iu1f7brY_normal.png
profile_image_url_https:
https://pbs.twimg.com/profile_images/451007105391022080/iu1f7brY_normal.png
profile_banner_url: https://pbs.twimg.com/profile_banners/813286/1360123769
profile_link_color: "2574AD"
profile_sidebar_border_color: "FFFFFF"
profile_sidebar_fill_color: "C2E0F6"
profile_text_color: "333333"

```



```

profile_use_background_image: true
default_profile: false
default_profile_image: false
following: false
follow_request_sent: false
notifications: false
}
geo: null
coordinates: null
place: null
contributors: null
retweet_count: 152569
favorite_count: 52730
-entities: {
  hashtags: [ ]
  symbols: [ ]
  urls: [ ]
  user_mentions: [ ]
}
favorited: false
retweeted: false
lang: "en"
}

```

Essència

Sempre podem dubtar de l'Essència, però mirarem llavors el codi JSON exposat i a veure si hi trobem res d'interessant.

Propietats Essencials

Fiabilitat

Que sigui fiable el text ens ho pot fer fiable la marca “-bo” que ens indica que la font prové d'Obama, o almenys així ens ho deia la capçalera del seu Twitter. I poca cosa més. Veiem que es publica en el seu perfil i per tant es fa fer seu.

Autenticitat

Val que la marca “-bo” és una mostra de la voluntat de voler distingir els missatges personals d'Obama de la resta de missatges del perfil de Twitter, però res més. Qualsevol podria ser el que fos qui escrigués els 3 caràcters, això sí amb l'aprovació d'Obama esperem.

Integritat

Integritat sí. No hi ha dubte, tenim tot el document complet, o almenys el que se'n va publicar a Twitter.

Si hi va haver una redacció anterior i després es va anar abreujant per a fer-la cabre en els 140 caràcters que requereix Twitter, això ja no ho sabem. Però tampoc sembla que s'hagués escurçat massa, ja que és prou breu i contenint 91 caràcters la idea quedava prou clara.

Identitat

Sabem que prové del perfil d'Obama perquè ens hi apareix el seu nom, el sobrenom i la marca blava de *verified account* del seu compte, com ja hem dit abans. El procés per obtenir una *verified*

account no és fàcil⁷⁶, i ens ofereix bastanta rigurositat, així que la identitat seria bastant clara.

Usabilitat

Veiem milers de retuits i favorits fets al tuit (`retweet_count: 152569 favorite_count: 52730`), així que podem creure que el document és usable. A més la quantitat de respostes en diversos idiomes ens permet veure com fins i tot des d'indrets de tot el planeta es van sentir cridats pel document.

Accessibilitat

No va ser difícil trobar-lo ja que s'havia publicat uns dos dies després de les votacions i resultats de les eleccions. Obama guanya el dia 5 de novembre, és nomenat president el 6 de novembre, i el tuit és enviat el 7. Si no, eines com Snapbird també ens permetrien trobar-lo fàcilment⁷⁷, ja que el cercador web de Twitter no arriba a nivells massa profunds a vegades i es podria arribar a haver de fer servir fins i tot la API també per a les cerques més profundes.

A més, si guardes el número ID del tuit després ja pots fàcilment mostrar-lo si saps quin esquema segueix Twitter amb les URL: `http://www.twitter.com/usuari/status/iddeltuit`.

Propietats legals

Tot aquest apartat de les propietats legals l'hem desenvolupat extensament ja en el sistema jurídic, executiu i judicial en el punt del Sistema jurídic d'aquest treball. Però si no, podem mirar ara què ens diuen les dades JSON i veure algun detall legal com per exemple si la marca “-bo” és fiable o esbrinar des d'on s'ha enviat i per tant si a més de la legislació americana que contemplava Twitter n'hauríem d'afegir alguna més.

La llàstima és que la geolocalització està desactivada segons ens informa JSON, però ens queda un petit detall desapercebut al principi però que hem vist al final: la part `"utc_offset" : -14400,`. Aclarirem que està expressat en segons, per tant parlem de -4 hores respecte el meridià de Greenwich, és a dir que hauria de correspondre a un aparell amb l'hora configurada per l'est del continent americà. Així que podríem considerar que s'ha enviat des d'algun indret d'allà. Però no sabem ni amb quin aparell ja que es va enviar a través de la web de Twitter `"source" : "web",`, així que juntament amb la geolocalització desactivada ens és poc possible saber qui va emetre finalment aquest document.

Podem dir doncs que respecte les propietats legals, el més probable definitivament és que la legislació que actuï sigui la dels EUA, la pròpia que ja tenia Twitter i prou. Acabant aquí les propietats legals segurament.

76 Twitter Help Center | FAQs about Verified Accounts <https://support.twitter.com/groups/31-twitter-basics/topics/111-features/articles/119135-about-verified-accounts> (consultat a 16 d'abril de 2014)

77 Snap Bird - search twitter's history <http://snapbird.org/> <http://snapbird.org/BarackObama/timeline/We%27re%20all%20in%20this%20together> (consultat a 16 d'abril de 2014)

Explicar més coses amb les metadades JSON

Òbviament a través de l'API de Twitter es poden extreure més metadades que les que acompanyen un sol tuit. La varietat de metadades no és tant completa com desitjaríem però si un fa prou voltes pot enginyar-se-les per extreure'n les més bàsiques. I si un se les enginya encara més pot arribar a treure'n sobre tendències, fluctuacions, estadístiques i més.

Però com dèiem en l'inici del treball quan parlàvem del plantejament que volíem fer, no ens posarem a exposar dades en gràfiques ni càlculs especulatius sobre el comportament dels usuaris de Twitter, sinó que pretenem tan sols analitzar quines dades i metadades ens ofereix la documentació generada a Twitter.

Com bé explica Twitter⁷⁸, cada tuit publicat és identificat amb un número de sèrie generat pel servei Snowflake que no és només un número seqüencial com les IDs dels usuaris, sinó basat també en el temps: El número final està compostat per una marca de temps, un número de "treball", i un número seqüencial. I així, després d'obtenir aquest número identificatiu, vindrà la part d'assignar les dades i metadades al tuit.

Els tuits són coneguts també com a "actualitzacions d'estat", i poden ser incrustats en webs, resposos, guardats a favorits o no, i eliminats. També poden ser reenviats a través del sistema "retuit" que els reenvia als seguidors de l'usuari que ha exercit aquesta acció.

Així que com ja hem exposat, a part de les dades i opcions visibles, de fons hi ha metadades que poden també ser incloses dins d'un tuit, tot i que no són sempre obligatòries ni estan activades totes per defecte. Entre les que anomena Twitter podríem trobar⁷⁹:

annotations: Encara en desús tot i que s'espera que en un futur aparegui en la versió Beta de Twitter, com a anotacions al propi tuit.

contributors: És la col·lecció de contribuïdors. Una metadada que es pot desactivar i que permetrà indicar quins usuaris han contribuït a l'autoria del tuit. També és una metadada que es troba en proves a la versió Beta i que no se'n sap massa cosa encara. Així que apareix en l'esquema de metadades però com a desactivada `contributors_enabled: false`.

```
"contributors":
[
  {
    "id":819797,
    "id_str":"819797",
    "screen_name":"episod"
  }
]
```

78 "Twitter IDs, JSON and Snowflake Twitter Developers." <https://dev.twitter.com/docs/twitter-ids-json-and-snowflake> (consultada a 17 d'abril de 2014)

79 "Tweets." Twitter Developers. <https://dev.twitter.com/docs/platform-objects/tweets> (consultat a 17 d'abril de 2014)

coordinates: També és un camp desactivable. I fa referència a la geolocalització del tuit, en format geoJSON (primer la longitud i després la latitud).

```
"coordinates":
{
  "coordinates":
  [
    -75.14310264,
    40.05701649
  ],
  "type": "Point"
}
```

created_at: Marca el temps en horari UTC (+0000) de quan es va crear el tuit.

```
"created_at": "Wed Aug 27 13:08:45 +0000 2008"
```

current_user_retweet: Detalla la identificació dels qui han fet retuit a un tuit en concret. Es pot activar i desactivar segons les preferències dels usuaris de voler-s'hi mostrar o no.

```
"current_user_retweet": {
  "id": 26815871309,
  "id_str": "26815871309"
}
```

Entities: Són les entitats que acompanyen la informació del tuit, tals com etiquetes (*hashtags*), direccions web (*URLs*), i mencions (*user_mentions*). Dins d'aquí trobaríem entitats també com "media" que designarien una imatge (amb la seva ubicació i mida a més de número identificatiu), "url" que designaria una direcció web, User Mention que seria la menció a un usuari (i que ens mostraria la seva id),... Curiós que en el cas de la imatge s'adjunti una URL i no una codificació en BASE64 per exemple, però suposem que la càrrega de treball del servidor deu ser menor amb la URL⁸⁰.

```
"entities":
{
  "hashtags": [],
  "urls": [],
  "user_mentions": []
}
```

favorite_count: També desactivable. Exposa aproximadament quantes vegades s'ha afegit a favorits el tuit per part dels usuaris de Twitter.

```
"favorite_count": 1138
```

favorited: Indica si el tuit ha estat afegit a favorits per l'usuari autenticat.

⁸⁰ "Entities." Twitter Developers. <https://dev.twitter.com/docs/platform-objects/entities> (consultat a 3 de maig de 2014)

```
"favorited":true
```

filter_level: Indica el nivell del llenguatge utilitzat al tuit, segons si provoca molt de soroll o és molt apte per a la seva emissió a la xarxa.⁸¹

```
"filter_level": "medium"
```

id: El número identificatiu del tuit. Com ja hem explicat abans el seu procés de creació és una tasca del servei Snowflake. Està signat a 64 bits per al seu emmagatzematge segur.

```
"id":114749583439036416
```

id_str: La representació de l'identificador únic del tuit, en versió resumida del número identificatiu original de l'anterior metadada.

```
"id_str":"114749583439036416"
```

in_reply_to_screen_name: Indica a qui nom d'usuari s'ha respost.

```
"in_reply_to_screen_name":"twitterapi"
```

in_reply_to_status_id_str: El número identificatiu resumit del tuit al que s'ha respost. Si el tuit representat és una resposta contindrà la cadena del número identificatiu del tuit original.

```
"in_reply_to_status_id_str":"114749583439036416"
```

in_reply_to_user_id: El número identificatiu de l'usuari al que s'ha respost. Si el tuit representat és una resposta contindrà la cadena del número identificatiu del perfil de Twitter original. No sempre serà l'usuari directament mencionat en el tuit.

```
"in_reply_to_user_id":819797
```

in_reply_to_user_id_str: Si el tuit representat és una resposta contindrà la cadena del número identificatiu resumit del perfil de Twitter original. No sempre serà l'usuari directament mencionat en el tuit.

```
"in_reply_to_user_id_str":"819797"
```

lang: Representarà la identificació de l'idioma en BCP 47 que s'ha detectat en el text del tuit. Apareixerà "und" si no se n'ha detectat cap.

```
"lang": "en"
```

place: Indicarà a quina localització està relacionada el tuit, que no té perquè ser la present a l'hora de publicar-lo. Va a intenció de l'autor del tuit.

```
"place":
```

81 Roomann-Kurrik, Arne. "Introducing new metadata for Tweets | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2013/introducing-new-metadata-for-tweets> (consultat a 17 d'abril de 2014)

```
{
  "attributes": {},
  "bounding_box": {
    "coordinates": [
      [
        [-77.119759, 38.791645],
        [-76.909393, 38.791645],
        [-76.909393, 38.995548],
        [-77.119759, 38.995548]
      ],
      "type": "Polygon"
    ],
    "country": "United States",
    "country_code": "US",
    "full_name": "Washington, DC",
    "id": "01fbe706f872cb32",
    "name": "Washington",
    "place_type": "city",
    "url": "http://api.twitter.com/1/geo/id/01fbe706f872cb32.json"
  }
}
```

possibly_sensitive: Indica si la direcció web que apareix al tuit té contingut apte per a tots els públics o no. No és una garantia que s'hagi analitzat el contingut però serveix d'indicació prèvia si s'indica.

```
"possibly_sensitive": true
```

scopes: Parell de claus que indiquen la verificabilitat del contingut del tuit. Aquesta metadada és utilitzada pels tuits promocionals "Twitter's Promoted Products".

```
"scopes": {"followers": false}
```

retweet_count: Número de vegades que un tuit ha sigut retuitejat.

```
"retweet_count": 1585
```

retweeted: Indica si el tuit ha sigut retuitejat per usuaris autenticats.

```
"retweeted": false
```

retweeted_status: Representa el tuit retuitejat.

source: Mostra des d'on s'ha enviat el tuit. Si s'ha fet des de la web de Twitter mostrarà el valor "web", en canvi si s'ha fet des d'una aplicació mostrarà el nom públic d'aquella aplicació.

```
"source": "\u003Ca
href=\"http://itunes.apple.com/us/app/twitter/id409789998?mt=12\"
rel=\"nofollow\"\u003ETwitter for Mac\u003C/a\u003E"
```

text: cadena de text codificada en UTF-8 o missatge inclòs en el tuit i visible com a contingut públic.

```
"text": "Tweet Button, Follow Button, and Web Intents javascript now support SSL http://t.co/9fbA0oYy ^TS"
```

truncated: Indica si el contingut del tuit ha estat troncat. Això es donava amb els retuits manuals anys abans, però des que hi ha un sistema establert per a fer els retuits aquest valor està desactivat.

```
"truncated": true
```

user: L'usuari que ha publicat el tuit. I les metadades en JSON del seu perfil⁸².

```
"user": {"statuses_count": 3080, "favourites_count": 22, "protected": false,
"profile_text_color": "437792", "profile_image_url": "...", "name": "Twitter API",
"profile_sidebar_fill_color": "a9d9f1", "listed_count": 9252,
"following": true, "profile_background_tile": false, "utc_offset": -28800,
"description": "The Real Twitter API. I tweet about API changes, service issues and happily answer questions about Twitter and our API. Don't get an answer? It's on my website.", "location": "San Francisco, CA",
"contributors_enabled": true, "verified": true,
"profile_link_color": "0094C2", "followers_count": 665829,
"url": "http://dev.twitter.com", "default_profile": false,
"profile_sidebar_border_color": "0094C2", "screen_name": "twitterapi",
"default_profile_image": false, "notifications": false, "display_url": null,
"show_all_inline_media": false, "geo_enabled": true,
"profile_use_background_image": true, "friends_count": 32,
"id_str": "6253282", "entities": {"hashtags": [], "urls": [],
"user_mentions": []}, "expanded_url": null, "is_translator": false,
"lang": "en", "time_zone": "Pacific Time (US & Canada)",
"created_at": "Wed May 23 06:01:13 +0000 2007",
"profile_background_color": "e8f2f7", "id": 6253282,
"follow_request_sent": false, "profile_background_image_url_https": "...",
"profile_background_image_url": "...", "profile_image_url_https": "..."}

```

withheld_copyright: Indica que el contingut del tuit ha superat la supervisió i compleix amb la DMCA. Aquesta metadada com totes les de "withheld" encara estan en proves i no se sap ben bé com s'implantaràn finalment.⁸³

```
"withheld_copyright": true
```

withheld_in_countries: Mostra les dues lletres del país en el que se situa el tuit.

```
"withheld_in_countries": ["GR", "HK", "MY"]
```

withheld_scope: Indica si el contingut ha estat restringit a un "status" o a un "user".

```
"withheld_scope": "status"
```

Totes aquestes metadades però poden anar ampliant-se, degut als constants canvis i

⁸² Com hem avisat en una anterior nota al peu, als Annexos del treball a "Manual per extreure dades en JSON amb Postman i l'API de Twitter" donem les indicacions i el manual per obtenir aquestes metadades en esquema JSON a través d'interactuar amb l'API de Twitter.

⁸³ Costa, Jason. "New Withheld Content Fields in API Responses | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2012/new-withheld-content-fields-api-responses> (consultat a 17 d'abril de 2014)

actualitzacions del sistema que proposa Twitter. En concret no hem trobat encara metadades que parlin de Vine, el servei de vídeos breus (6 segons o menys) que es va estrenar al gener de 2013⁸⁴. Tampoc n'hem vist sobre les noves opcions d'etiquetar usuaris a les fotografies adjuntades als tuits, ni tampoc sobre adjuntar vàries fotografies en un sol tuit⁸⁵. Suposem que ja aniran apareixent en quan es vagin popularitzant més.

2.4 L'estudi de Twitter

No tenim un corpus teòric ni molt menys pràctic de com tractaria un arxiver la documentació de Twitter, ja que l'únic projecte en ferm que hi ha és el de la Biblioteca del Congrés dels Estats Units i encara es troba en estudi intern, per tant només ens queda especular sobre el que s'ha publicat per sobre.

2.4.1 La credibilitat de Twitter

El principal entrebanc de Twitter és quina credibilitat li donem al que s'hi publica. De fet com ja hem dit es tracta d'una xarxa de microblogs personals on cada usuari hi penja el que li ve de gust i com li ve de gust⁸⁶. Però el món acadèmic ja s'ha animat a investigar-ho en profunditat i es poden trobar estudis científics que analitzen fins a quin punt podem considerar la fiabilitat que ens ofereix Twitter.

L'estudi de Carlos Castillo, Marcelo Mendoza, i Barbara Poblete, del món acadèmic xilè, s'ha plantejat el tema de com la rumorologia s'expandeix a través de Twitter, degut a la possibilitat de reenviar i reenviar els tuits a gran velocitat. En la seva publicació "Information Credibility on Twitter"⁸⁷ plantegen com Twitter permet estar alertats de històries puntuals en un punt geogràfic, o poder geolocalitzar esdeveniments. De fet anomenen estudis en els que s'ha constatat ja des del 2008 que la principal font des d'on la població extreu les notícies sol ser Internet, malgrat que els blogs són vistos amb menys confiança que les pàgines de notícies convencionals. Tampoc hi ha ajudat el fet que sovint es desafiïn sistemàticament els sistemes de seguretat i s'hagin donat alguns casos d'intrusió en comptes oficials de dirigents polítics o personalitats públiques. També s'ha intentat establir patrons automatitzats sobre com detectar la fiabilitat del que es publica, amb

84 Sippey, Michael. "Vine: A new way to share video | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2013/vine-a-new-way-to-share-video> (consultat a 17 d'abril de 2014)

85 Mateos, Manu. "Hasta cuatro fotografías por tweet y además etiquetadas: novedades de un Twitter más social." Hasta cuatro fotografías por tweet y además etiquetadas: novedades de un Twitter más social. <http://www.genbeta.com/redes-sociales-y-comunidades/twitter-cada-vez-mas-como-facebook-etiquetas-en-fotos-y-hasta-cuatro-en-cada-tweet> (consultat a 17 d'abril de 2014)

86 De fet en els inicis, del 2007 al 2012 hi havia certa competició sobre qui aconseguia enviar tuits des de la plataforma o aparell més inversemblant, ja fos la nevera intel·ligent, o el terminal del telèfon mòbil més rudimentari. Tot això aniria minvant a mesura que els requisits de Twitter anaven augmentant, la API s'anava restringint, i apareixien elements multimèdia com fotografies o vídeos adherits als tuits que exigien ja disposar d'una pantalla prou adaptada als usos nous que anaven sorgint.

87 Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. Proceedings of the 20th international conference on World wide web - WWW '11: 675. <http://portal.acm.org/citation.cfm?doid=1963405.1963500> (consultat a 18 d'abril de 2014)

paràmetres com el vocabulari positiu en els tuits de resposta o si se citen noves webs que aportin més informació a la notícia, també el número de seguidors dels usuaris que estan reenviant la notícia⁸⁸, i tot de paràmetres respecte si l'usuari té la merca de verificat, si els seus tuits són molt extensos, si té posada una descripció personal al seu perfil,... entre d'altres. Però les conclusions són que com més urgència tenen els usuaris de Twitter, més fàcilment donen credibilitat a qualsevol informació, i si s'inclou alguna direcció web les probabilitats augmenten. Com que l'estudi és de la primavera 2011 segurament encara no es podien incloure fotografies als tuits ja que aquesta característica no es va introduir fins a l'estiu del 2011⁸⁹, per tant no sabem si el poder incloure encara faria incrementar més les possibilitats, o si incloure més elements multimèdia adjunts al tuit característica que es presentà al juny de 2012⁹⁰ encara les incrementaria molt més.

En un altre estudi anomenat "Evaluating Event Credibility on Twitter" publicat per Manish Gupta, Peixiang Zhao i Jiawei Han de la University of Illinois⁹¹, s'analitzà al 2012 la influència de Twitter a través també de quina credibilitat s'hi donava a les notícies publicades a través d'aquest mitjà. Destacaven que un 4% del que es publicava a Twitter eren directament notícies. Tinguem en compte però que un 44% dels comptes a Twitter no ha publicat mai cap tuit⁹², i que tot i que un 8% aproximadament donava credibilitat a Twitter, eren un 12% aproximadament els qui donaven credibilitat al que publicaven els seus amics a Twitter. L'estudi aquest feia una crítica a l'anterior estudi que hem esmentat de Castillo, Mendoza i Poblete, ja que segons diuen analitzaven molt el fenomen de l'esdeveniment, però s'oblidaven de la relació entre tuits i usuaris, entre el missatge que es dona i qui el rep, de l'empatia personal que es poguessin tenir emissor i receptor. Per això no van analitzar tuit per tuit, sinó que van englobar per apartats: anàlisi de les varietats de tuits segons l'esdeveniment, classificació d'usuaris segons les seves interaccions, números d'actualitzacions, amics (gent a qui segueix i que ell també segueix), antiguitat de l'usuari a Twitter, vocabulari informal utilitzat en els tuits, manera d'incloure les direccions web, si inclou localitzacions del que explica als tuits, si menciona a d'altres usuaris i si són o no seguidors seus o amics,... i si interactuava amb d'altres usuaris de Twitter dels que se n'hagués reconegut la credibilitat. A més, veien la problemàtica de la ironia i dels dobles sentits, sobretot quan es feien retuits o quan s'afegia algun tuit a favorites, ja que podia haver-hi la interacció d'una opinió personal

88 Aquest aspecte pot ser perillós ja que és més fiable buscar la relació seguits/seguïdors per tal d'establir la rellevància i interès que desperta l'usuari concret de Twitter.

89 Melanson, Donald. "Twitter announces integrated photo-sharing service, improved search." Engadget. <http://www.engadget.com/2011/06/01/twitter-announces-integrated-photo-sharing-service-improved-sea/> (consultat a 18 d'abril de 2014)

90 Sippey, Michael. "Experience more with expanded Tweets | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2012/experience-more-with-expanded-tweets> (consultat a 18 d'abril de 2014)

91 Gupta, Manish, and Peixiang Zhao. 2012. Evaluating Event Credibility on Twitter. In Proceedings of SIAM International Conference on Data Mining, 153-164. http://www.cs.uiuc.edu/~hanj/pdf/sdm12_mgupta.pdf (consultat a 18 d'abril de 2014)

92 López, Miguel. "El 44% de las cuentas de Twitter jamás ha enviado ningún tuit." El 44% de las cuentas de Twitter jamás ha enviado ningún tuit. <http://www.genbeta.com/redes-sociales-y-comunidades/el-44-de-las-cuentas-de-twitter-jamas-ha-enviado-ningun-tuit> (consultat a 18 d'abril de 2014)

del propi usuari de Twitter que en principi podríem desconèixer completament. Un aspecte que podia ajudar a destacar la importància i per tant indicar-nos on hi podia haver credibilitat era amb la concentració espontània de tuits en una temàtica, tot i que els Trending Topics tampoc eren cap base totalment fiable ja que com és sabut dins de Twitter es programen campanyes massives de publicar en un mateix moment tants tuits com es puguin contenint una mateixa etiqueta (*hashtag*) per tal de fer-se veure a mode de propaganda a l'apartat de Trending Topics i cridar l'atenció pública per un interès propagandístic. Tot plegat ho van presentar en un projecte que anomenaren Truthy⁹³ i que estudia com entendre les difusions i comunicacions de temes polítics i moviments socials a Twitter. Aquest projecte és capaç de mostrar-nos dades prou bàsiques i interessants com la freqüència d'ús de Twitter que fa @barackobama⁹⁴. Les seves conclusions és que no es podia analitzar només la banda de qui publica els tuits i que s'havien de fer anàlisis creuats amb els esdeveniments concrets, les relacions dels usuaris, i els tuits que es publicaven.

Al març de 2012 va aparèixer un nou estudi titulat "Credibility Ranking of Tweets during High Impact Events"⁹⁵ en el que Aditi Gupta i Ponnurangam Kumaraguru de l'Indraprastha Institute of Information Technology a Delhi (la Índia), tornaven a desenvolupar el tema de la credibilitat dels missatges a Twitter. Aquest estudi tenia l'interès que es mostrava molt conscient del paper de Twitter com a canal d'informació que s'alimentava de la informació rebuda re-transmetent-la a qui li interessava seguir-la. És per això que van dedicar-se a analitzar diversos esdeveniments i què se'n deia a Twitter com l'Huracà Irene. Descobrint que un 30% dels missatges informaven de l'esdeveniment i un 14% eren bàsicament spam, però tan sols un 17% del contingut era fiable segons els usuaris. Segons aquest estudi un 85% del que es publica a Twitter són notícies, i que tot i no mostrar-se un entusiasme gaire elevat per les notícies internacionals sí que hi havia certa activitat en difondre-les. Entre les indicacions més destacades estaven que com més preguntes apareixien en els tuits més rumors estaven circulant sobre aquella notícia, també que com més rellevància agafava una notícia més direccions web apareixien als tuits i menys mencions (@nomusuari) apareixien. L'estudi a més va crear tota una taula amb les diverses notícies rellevants triades, a les quals van assignar etiquetes fiables (els *hashtags* més usats), enllaços a notícies més fiables; i així van anar classificant els diversos tuits. I van acabar conclouent que no importava massa qui era el que publicava el tuit sinó quina informació aportava. És l'empatia humana la que estableix el grau de confiança i el que és més veritable.

Malgrat tot aquest reguitzell de complicacions la intenció d'aprofitar Twitter per part dels

93 "Truthy Information diffusion research at Indiana University." Truthy. <http://truthy.indiana.edu> (consultat a 18 d'abril de 2014).

94 "@barackobama 835,402 tweets in the past 9 months This meme is TruthySpamTweet." @barackobama Truthy. http://truthy.indiana.edu/memedetail?id=783&theme_id=4 (consultat a 18 d'abril de 2014)

95 Gupta, Aditi, and Ponnurangam Kumaraguru. "Credibility ranking of tweets during high impact events." Credibility ranking of tweets during high impact events. <http://dl.acm.org/citation.cfm?id=2185356> (consultat a 2 de maig de 2014)

arxivers s'ha destacat també en certs moments, com veurem en el següent apartat, tot i que de manera puntual i no a un nivell tant profund d'anàlisi tècnic i sistemàtic com en els dos estudis que acabem d'esmentar.

2.4.2 Implicació a Twitter

Al maig de 1997 veiem com Ferran Agelet parla que hi ha una consciència que Internet és un espai on l'arxivística s'hi ha d'implicar, tant a nivell intern com extern.⁹⁶ Ja es parlava del Directori de pàgines web d'arxius de l'Associació d'Arxivers de Catalunya, basat en el treball de Betlem Martínez Raduà, o del directori "Áncora" realitzat per estudiants de la Universitat Carlos III de Madrid amb llavors 500 adreces. Avui dia però la realització dels directoris ha disminuït la seva urgent necessitat degut a la indexació contínua que fa el cercador Google. De l'article del Ferran però ens quedem amb els seus "10 motius que justifiquen la connexió a Internet d'un arxiu": ... "2. Fer valer els coneixements dels arxivers, quant a la gestió del patrimoni documental, en la societat de la informació. Especialment l'avaluació documental"... "6. Participar en el mercat de la informació que s'està construint a Europa" ... "8. Participar en projectes interdisciplinaris de Recerca i Desenvolupament" ... "9. Desenvolupar la professió per tal d'obrir nous horitzons a l'arxiver" ... "10. Participar en el disseny i la creació dels sistemes i dels serveis d'informació.". Ja que al món informàtic el que estem explicant en aquest treball sobre Twitter els pot semblar molt banal, però per als arxivers pot suposar un pas més cap a implicar-se en nous horitzons, com s'ha animat a fer la Llibreria del Congrés dels Estats Units.

Per la seva banda Jordi Graells i Costa, parla de com les administracions no han de «veure les xarxes socials no com unes plataformes estranyes, ni com un mitjà ideal per aconseguir un creixement ràpid de l'audiència, sinó com unes eines aliades»⁹⁷. Aquells instruments que serviran per compartir, participar, escoltar i conversar.

De l'article de Jordi Graells, cal destacar la informació que ens dona sobre que l'obertura de nous perfils a les xarxes socials per part dels arxius, s'ha de notificar al Departament de la Presidència, la qual certifica aquesta obertura. I n'autoritza el compte per part de la DG d'Atenció a la Ciutadania i Difusió, atorgant-ne el nom més indicat i aplicant la imatge corporativa fixada en el PIV (Programa d'identificació visual) a mode de certificació. Pel tema del nom ho trobem correcte ja que s'ha d'establir certa coherència nominal. Pel tema de que la imatge corporativa sigui només l'avatar és molt precari ja que en el tema de Twitter com ja hem dit és imprescindible comptar almenys amb la marca de *verified account* per a certificar validesa d'identitat del compte.

Així doncs ens hem disposat a revisar aquesta "Guia d'usos i estil a les xarxes socials de la

96 AGELET, F. (1998). "Internet, un nou horitzó per als arxius. L'experiència de l'Arxiu Municipal de Barcelona". A: Lligall [Barcelona], núm. 13, p. 185-204.

97 GRAELLS, J. (2011). "Les xarxes socials revolucionaran la professió d'arxiver". A: Lligall [Barcelona], núm. 32, p. 157 – 171.

Generalitat de Catalunya”⁹⁸ que citava Jordi Graells per veure què se’n deia de Twitter. I hem vist que aquesta és una guia elaborada al 2010, en la que s’exposen recomanacions per a introduir-se en les xarxes socials des dels organismes públics de la Generalitat de Catalunya. En el cas de Twitter s’hi donen indicacions com que «La denominació dels comptes de departaments o serveis es forma a partir del nom propi de l’àmbit, servei o marca i afegint-hi el sufix “cat”. Per exemple: Justícia @justiciacat, Salut @salutcat, Emergències @emergenciescat.»⁹⁹. El que se’ns fa més confús però és l’administració d’aquests comptes, ja que es recomana «evitar treballar directament amb l’identificador i la contrasenya de les diferents aplicacions de xarxa social». Desconeixem el significat d’aquesta afirmació, com també la lògica que fer més farragosa l’administració l’ha de fer més segura. A més recomanen crear comptes a l’aplicació Hootsuite per gestionar Twitter, una externalització que a no ser que es busquin funcions extra que no es troben a Twitter (programar tuits, seguiment de varis hashtags,...) no entenem. La DGAC a més designa un usuari avançat que fa d’administrador, i aquesta persona nomenarà més usuaris editors corporatius. Penso que no faltarà massa perquè Twitter faciliti aquesta funció ja que com hem vist la metadada “contributors” es té en compte ja i segurament s’activarà quan Twitter ho cregui oportú, però mentrestant doncs hi ha el que hi ha. Per a comprovar si es complia aquesta recomanació oficial, hem volgut i pogut comprovar-ho a través de la API¹⁰⁰ que tant el compte @govern (<http://twitter.com/govern>) com @gencat (<http://twitter.com/gencat>) no fan servir Hootsuite sinó Tweetdeck, l’aplicació avançada que va adquirir ja fa temps Twitter¹⁰¹ i que es considera per tant la oficial i més recomanable¹⁰².

A més de la direcció web del departament, servei o marca, la localització ha de ser Catalunya. Una descripció (Bio) breu que ja es designa com a “Projectes i activitat de [nom del departament, servei o marca] de la Generalitat de Catalunya” i que no ha de tenir més de 160 caràcters. A més es marquen uns colors en hexadecimal de disseny del perfil, fons #ffffff, text #333333, enllaços #ac2215, barra lateral #ffffff i vora de la barra lateral #4d4d4d. De fons una imatge composta per la marca general del departament en qüestió i la URL pròpia del departament, servei o marca en format apaïsat.

98 Guia d’usos i estil a les xarxes socials de la Generalitat de Catalunya http://www.gencat.cat/xarxessocials/pdf/v1_guia_usos_xarxa_cat.pdf (consultada a 23 d’abril de 2014)

99 Guia d’usos i estil a les xarxes socials de la Generalitat de Catalunya http://www.gencat.cat/xarxessocials/pdf/v1_guia_usos_xarxa_cat.pdf pg. 13 (consultada a 23 d’abril de 2014)

100 Als Annexos a “Manual per extreure dades en JSON amb Postman i l’API de Twitter → Les metadades d’un perfil” explicarem com extreure les metadades de tuits concrets d’aquests comptes.

101 Falconer, Joel. "Deal Done: Twitter acquires TweetDeck for over \$40m - The Next Web." TNW Network All Stories RSS. <http://thenextweb.com/twitter/2011/05/24/deal-done-twitter-acquires-tweetdeck-for-over-40m/> (consultat a 23 d’abril de 2014)

102 Malgrat els greus problemes soferts recentment: Hern, Alex. "Twitter shuts down Tweetdeck after XSS flaw leaves users vulnerable to account hijack." theguardian.com. <http://www.theguardian.com/technology/2014/jun/11/twitter-tweetdeck-xss-flaw-users-vulnerable> (consultat a 15 de juny de 2014)

Com a idioma es recomana que «els tuits es redacten majoritàriament en català, la llengua comuna i pròpia», tot i que també es permet el castellà i l'anglès, «segons el tipus de continguts de què es tracti». I «pel que fa als retuits, es respectarà la llengua en què han estat escrits originàriament», una obvietat ja que com a norma no es tradueixen mai els retuits¹⁰³, si no deixen de ser retuits per a ser reenviaments de tuits. A més, la recomanació d'activitat és d'1 a uns 10 tuits diaris per mantenir certa continuïtat i normalitat, excepte en casos que es retransmeti un acte en directe o s'informi sobre emergències.

És interessant com fins i tot es fa front a com s'han de gestionar les respostes a les crítiques, sempre de forma manual per ser el més amables i personals possible. I en el cas que siguin negatives s'ha de fer una reflexió més profunda i extensa per poder respondre amb la millor informació i el més acurada possible de forma constructiva. A més per a requerir més informació no es recomanen els missatges privats sinó el correu electrònic, cosa que pot ser un encert en casos molt complexos. També és remarcable que es faci ús d'un dipòsit de respostes en format wiki al portal d'Atenció Ciutadana, grup Internet, de la plataforma e-Catalunya, «perquè puguin servir per a altres ocasions a tota la comunitat que gestiona la presència de la Generalitat a les xarxes socials».

En l'apartat "3.6 Verificar el compte", s'ofereix alguns consells sobre com verificar el compte oficial a Twitter i com omplir el formulari de verificació. Aquest mètode ja no està en funcionament, i ara és directament posant-nos en contacte amb Twitter i tenint el recolzament d'algun organisme reconegut públicament que podrem verificar-nos el compte, sempre i quan siguem una personalitat o institució pública rellevant.

Aquesta guia que hem vist és molt similar a nivell internacional a la d'altres organismes com per exemple la de "Twitter Guidelines and Best Practices"¹⁰⁴ de la agència Centers for Disease Control nord-americana. Tot i que aquesta guia estableix fins i tot que en els informes s'han de registrar els missatges de Twitter més rellevants, les respostes rebudes a través d'aquesta xarxa social, per tal de complir amb les indicacions del registre i arxiu federals. Així com també tenir cura de com millorar la repercussió de les comunicacions i avaluar l'impacte d'aquestes, alhora que estar alerta de a quines campanyes de la xarxa social es pot col·laborar i participar per tal de ser més propers als ciutadans.

També és similar a la "Guía de usos y estilo en las Redes Sociales del Gobierno Vasco"¹⁰⁵, que a

103 Twitter ja ha implantat un sistema de traducció automatitzada de tuits: Edward, Rivas. "Twitter expande traducciones impulsados por Bing." *Tecnología geek*. <http://tecnologiageek.com/twitter-expande-traduccion-impulsados-por-bing-para-ios-la-web/> (consultat a 15 de juny de 2014)

104 Centers for Disease Control and Prevention. "Twitter Guidelines & Best Practices." Centers for Disease Control and Prevention. <http://www.cdc.gov/socialmedia/tools/guidelines/twitter.html> (consultat a 3 de maig de 2014)

105 "Guía de usos y estilo en las Redes Sociales del Gobierno Vasco." CAST.indd. http://www.irekia.euskadi.net/assets/a_documents/1218/Gui%CC%81a_de_usos_y_estilo_en_las_Red_Sociales_del_Gobierno_Vasco.pdf (consultat a 6 de maig de 2014)

més en el seu cas específic estableix protocols d'ús de l'idioma euskera a Twitter seguint la "Ley 10/1982, de 24 de noviembre, Básica de Normalización del Uso del Euskera". A Twitter prioritzen la llengua basca fins al punt que els *hashtags* (etiquetes temàtiques de Twitter) es recomana fer-los en basc ja que són més sintètics i breus. En el cas de la Generalitat de Catalunya els comptes de Twitter s'havien de fer des d'un correu de *gencat.cat*, en el del Govern Basc es fan des de correus *ej-gv.es*. En el seu cas recomanen gestionar els comptes amb CoTweet¹⁰⁶, no entenc massa perquè ni tampoc s'argumenta a la guia més enllà de que té capacitat de gestionar varis comptes alhora. Sorprenent em sembla quan aquesta eina si es visita el seu perfil a Twitter porta quasi 2 anys sense donar senyals d'activitat. Al igual que en el cas català, en el cas basc també es donen indicacions sobre com personalitzar els perfils o la quantitat recomanada de tuits a publicar al dia (totes coincideixen en d'1 a 10 diaris). Es recomana molt l'ús de Twitter perquè permet conèixer de primera mà i per una via molt senzilla quines són les consultes més habituals, és per això que la guia del Govern Basc recomana també anotar aquelles consultes més repetides i les respostes donades en un entorn wiki per tal d'anar construint així una eina que després els informàtics de la pàgina web de la institució puguin fer servir de model per a l'apartat Preguntes Més Freqüents de la web institucional.

Però tornem altra vegada als estudis arxivístics. Un altre estudi que va voler comprovar la implicació dels arxius a Twitter va ser el d'Enric Terradellas Prat, el qual a "L'ús de twitter en el món dels arxius"¹⁰⁷ es proposava fer una mostra de en quina situació es trobaven els arxius en la seva incursió a Twitter. L'estudi encara buscava determinar la identitat dels perfils de Twitter a través d'en quin idioma escriuen els tuits, no pas cercant amb la API quines metadades estaven associades als perfils¹⁰⁸, ja que amb una discriminació dels resultats cercant "lang": "ca" hagués identificat més fàcilment els perfils en català¹⁰⁹.

També analitzà qui publicava més a Twitter, però no mostrava qui rep més visites als enllaços inclosos als tuits, tot i que diu que Hootsuite donaria aquesta dada. Dada que sovint hi ha rumors que Twitter maneja i que podria aparèixer en un índex numèric en les metadades dels tuits, però que encara no se n'ha vist cap rastre a la API. En l'article l'Enric Terradellas parla sovint de Klout¹¹⁰, una eina que fa anàlisi i calcula a través de variables (seguidors, publicacions, respostes,...) quin nivell s'ha de determinar a cada usuari per tal de servir de referent als qui

2014)

106 CoTweet (cotweet) on Twitter <https://twitter.com/cotweet> (consultat a 11 de maig de 2014)

107 TERRADELLAS PRAT, Enric. "L'ús de twitter en el món dels arxius." Lligall 33 i 34. Revista catalana d'arxivística 2011-2012. Miscel·lània. Associació d'Arxivers i Gestors de Documents de Catalunya pg. 36-37

108 TERRADELLAS PRAT, Enric. "L'ús de twitter en el món dels arxius." Lligall 33 i 34. Revista catalana d'arxivística 2011-2012. Miscel·lània. Associació d'Arxivers i Gestors de Documents de Catalunya pg. 42

109 Fent una cerca a través de la API segons les indicacions que dona la pàgina de Twitter a: GET search/tweets | Twitter Developers <https://dev.twitter.com/docs/api/1.1/get/search/tweets>

110 TERRADELLAS PRAT, Enric. "L'ús de twitter en el món dels arxius." Lligall 33 i 34. Revista catalana d'arxivística 2011-2012. Miscel·lània. Associació d'Arxivers i Gestors de Documents de Catalunya pg. 43

busquen els comptes més influents de Twitter. Pot ser una eina interessant per a qui li interessi fer "carrera mediàtica" a Twitter, però serà enganyós creure que la influència es basa en qui sap difondre més només, ja que el 90% dels continguts compartits a Twitter estan generats pel 10% dels usuaris¹¹¹.

És interessant el treball de l'Enric per com exposa amb números el gran seguiment que tenen els comptes de Twitter dels centres arxivístics anglosaxons¹¹². També per com analitza quina activitat tenen els arxius catalans a Twitter i com estan utilitzant-ho per difondre'n els continguts que els semblen interessants i atraure als seus usuaris. I la categorització del tipus de perfil de Twitter amb el tipus de continguts que comparteix: arxius (promoció de continguts propis), associacions d'arxivers (difusió de continguts i actes interessants per als seus socis i seguidors), grups i col·lectius interessats en l'arxivística (promoció de continguts que poden resultar rellevants per al conjunt de la professió).

2.4.3 Els diaris personals i Twitter

Com ja hem vist, dins del món acadèmic s'han plantejat molts estudis sobre Twitter. Des dels que plantejaven aspectes més numèrics fins als que ho feien en temes més socials. Però també en trobem alguns com "Historicizing New Media: A Content Analysis of Twitter"¹¹³, que van per la vessant més social. L'estudi d'aquest article va recollir uns 101.069 tuits a analitzar. I en concret va voler determinar com Twitter es converteix en un reflex diari dels pensaments i activitats vitals de bona part de la població.

Mentre que avui dia es creu que els "diaris" personals s'omplen amb pensaments molt íntims que es volen mantenir en secret, no sempre ha sigut així. Al llarg de la història s'han redactat diaris personals pensats per a ser compartits i ensenyar bones lliçons als qui els havien de llegir. Ja al segle XVII fins al XIX els diaris van prendre un caràcter semi-públic, i sobretot els fets per dones eren sovint difosos per tal de llegar els coneixements a la família i afavorir les relacions socials, a l'estil dels *Speculum* o els *Exemplum* medievals, i demés literatura moral i tractats pedagògics com el "*Liber manualis Dhuodane quem ad filium suum transmisit Wilhelmum*" (malgrat que aquesta literatura antiga no sigui estrictament en format de diaris personals). Aquests diaris personals fins i tot s'enviaven i servien de diaris de viatge per tal de poder conservar les experiències i vivències al tornar a casa, i es compartien amb els amics com fa uns anys es compartien els àlbums de fotografies dels viatges, o com ara es comparteixen en comunitats

111 CBC/Radio Canada. "10% of Twitter users generate over 90% of content, study finds - Technology & Science - CBC News." CBCnews. <http://www.cbc.ca/news/technology/10-of-twitter-users-generate-over-90-of-content-study-finds-1.791107> (consultat a 6 de maig de 2014)

112 TERRADELLAS PRAT, Enric. "L'ús de twitter en el món dels arxius." Lligall 33 i 34. Revista catalana d'arxivística 2011-2012. Miscel·lània. Associació d'Arxivers i Gestors de Documents de Catalunya pg. 53

113 Historicizing New Media: A Content Analysis of Twitter
<http://onlinelibrary.wiley.com/doi/10.1111/jcom.12030/abstract;jsessionid=1D0DEDEBB16234374A9486E98947D048.f01t03> [<http://onlinelibrary.wiley.com/doi/10.1111/jcom.12030/pdf>]
 (consultat a 11 de maig de 2014)

virtuals tot tipus de documents que relaten els moments més importants dels viatges. No va ser comú introduir mecanismes de bloqueig en aquests diaris fins a 1860.

El que volem explicar és que Twitter ha oferit aquest espai públic des d'on compartir amb altres persones els continguts que un desitja difondre sobre sí mateix, el micro-blogging podria ser pres com uns nous diaris personals que van cap a la direcció altra vegada de ser més públics que privats. A més dins d'aquesta comparació entre els diaris personals i Twitter cal pensar que els diaris personals tampoc tenien una mida massa gran i moltes vegades les frases havien de ser molt breus i concretes. També podem comparar la temàtica del que s'hi escrivia ja que tant en aquells diaris personals com a Twitter s'hi podria trobar fets d'allò més quotidians com alhora relacions de fets més internacionals. Tal com ho defineix Zeynep Arda¹¹⁴, una de les gràcies de Twitter és que permet enviar un missatge personal a una audiència anònima i alhora a gent coneguda. Proveeix a una gran massa de població d'un mitjà massiu de comunicació.

2.4.4 Històries amb Twitter

Dins del món de la divulgació també trobem projectes més estrictament pedagògic i orientats al món de la divulgació històrica i cultural. Podríem parlar de com els diversos museus i arxius del món han fet ús de Twitter per divulgar la seva imatge i productes, però això ja ha estat explicat en varis articles com el que hem citat abans de d'Enric Terradellas "L'ús de twitter en el món dels arxius" per exemple.

El cas de TwHistory¹¹⁵ és el d'un projecte iniciat per difondre el coneixement de la Història a través de Twitter. La idea és que tu mateix et facis un dossier ben documentat sobre un gran esdeveniment, anotant el màxim d'esdeveniments dia per dia, i un cop tinguis omplert un període de temps prou ampli et registris a TwHistory i comencis a configurar personatges (que estaran representats en comptes de Twitter) que aniran interactuant per tal de fer una recreació en temps real dels fets diaris que tenies tu documentats sobre un gran esdeveniment¹¹⁶. Per a documentar-se suggereixen prendre com a referències els diaris, cartes i d'altra documentació cronista de l'època i els episodis concrets. Com a exemples ja desenvolupats se citen la batalla de Gettysburg o la Crisi dels Míssils a Cuba.

A Catalunya hem trobat un exemple força rellevant d'aquest tipus amb els perfils de Twitter creats per recrear la Guerra de Successió de 1714¹¹⁷, dirigit pel portal 3cat24 i que comptaren amb la

114 Arda, Zeynep . " Yo soy mi imagen 2.0: Aproximación a formas contemporáneas de construir la identidad humana." Tesis Doctorals en Xarxa. <http://www.tdx.cat/handle/10803/81855> (consultat a 11 de maig de 2014) pg. 259

115 Jensen, Marion, Caswell Tom, Ball Justin, Duffin Joel, and Barton Rob. "TwHistory : Sharing History Using Twitter." Repositori institucional - UOC. <http://openaccess.uoc.edu/webapps/o2/handle/10609/4942> [http://openaccess.uoc.edu/webapps/o2/bitstream/10609/4942/6/Caswell_editat.pdf] (consultat a 11 de maig de 2014)

116 "About." TwHistory. <http://twhistory.com/about> (consultat a 11 de maig de 2014)

117 "Especial Viu l' 11 de setembre de 1714 a Twitter." 3cat24. <http://www.324.cat/1714/> (consultat a 11 de maig de 2014)

col·laboració d'historiadors com Dani Cortijo¹¹⁸ i l'il·lustrador Francesc Riart, amb l'assessorament lingüístic d'Oriol Camps cap de lingüistes de la CCMA que avalà el respecte del llenguatge de l'època adaptat al parlar d'avui i al propi del mitjà¹¹⁹.

També des de la revista Sàpiens s'ha anat aprofitant la seva xarxa de blogs¹²⁰ per tenir tot un grup d'historiadors molt actius a Twitter. Però si hi ha un punt d'encontre és el del perfil d'historiadors a Twitter @historiadors20¹²¹, el qual ofereix i divulga els continguts que els historiadors catalans o d'altres nacionalitats publiquen a Twitter sota l'etiqueta #historiadors20.

3 Cos de la recerca: Casos arxivístics amb Twitter

Tal i com hem anunciat al final de l'apartat "2.2. Dificultats del treball", ens hem proposat també en aquest treball exposar alguns dels casos en que Twitter ha entrat de ple en el món de l'Arxivística. Així doncs ens disposem a veure alguns casos en que s'ha treballat o s'ha proposat treballar en termes arxivístics sobre Twitter.

3.1 L'exemple de NARA

El 14 d'abril de 2010 la Library of Congress¹²², la institució cultural més antiga del govern federal dels Estats Units i la biblioteca més gran del món¹²³, que té com a missió la recerca i la recepció de cada llibre, pamflet, mapa, impressió i peça de música registrada als Estats Units, va decidir també preservar tots els tuits públics que durant anys s'han anat creant a Twitter i que siguin dignes de preservació¹²⁴. Aquesta decisió es degué a que cada vegada Twitter està jugant un paper més rellevant en la història i política americanes des d'eleccions històriques fins a desastres devastadors. Aquests tuits doncs seran donats a la Library of Congress per a la seva preservació i recerca. Segons Twitter només hi haurà un retard de 6 mesos fins que la Library of Congress en disposi per a la recerca no comercial, i per a l'exposició pública, i preservació, i la data més antiga de la documentació que s'oferirà serà la del març de 2006 és a dir el dia 21 d'aquell mes que és

118 Cortijo, Dani. "Dani Cortijo (AltresBcn)." Twitter. <https://twitter.com/AltresBcn> (consultat a 11 de maig de 2014)

119 "El 3cat24 reviu l'11 de setembre de 1714 al Twitter." 3cat24. <http://www.324.cat/noticia/1363586/catalunya/El-3cat24-reviu-l11-de-setembre-de-1714-al-Twitter> (consultat a 11 de maig de 2014)

120 "Blogs – Sapiens.cat" <http://www.sapiens.cat/ca/blogs.php> (consultat a 14 de maig de 2014)

121 "#historiadors20 (historiadors20)." Twitter. <https://twitter.com/historiadors20> (consultat a 14 de maig de 2014)

122 Raymond, Matt. "How Tweet It Is!: Library Acquires Entire Twitter Archive." Library of Congress Blog. <http://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/> (consultat a 2 de maig de 2014).

123 La Biblioteca d'Alexandria encara no s'ha recuperat, mal ens pesi als habitants de la Mediterrània. 158.007.115 llibres té catalogats la Library of Congress (<http://www.loc.gov/about/general-information/> "General Information | Library of Congress" consultat a 3 de maig de 2014) i l'actual Biblioteca d'Alexandria està al nivell de 5.838.254 segons el seu "Annual Report 2011-2012" (http://www.bibalex.org/attachments_en/Publications/Files/2013011015545420497_annualreport2012sm.all.pdf pg. 66 (consultat a 3 de maig de 2014)

124 Stone, Biz. "Tweet Preservation | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2010/tweet-preservation> (consultat a 25 d'abril de 2014)

quan Jack Dorsey va publicar el seu primer tuit¹²⁵.

Així va quedar plasmat en el contracte de cessió ("GIFT AGREEMENT") que van firmar el conseller general de Twitter Alexander Macgillivray i James H. Billington de la Library of Congress¹²⁶. En el contracte Twitter feia donació als Estats Units d'Amèrica i més directament a la Library of Congress de la col·lecció consistent en els tuits públics de Twitter, a més de tot el material més directament relacionat amb aquests tuits i que fos imprescindible per a la seva obtenció i propòsit de preservació. Es deia en el contracte que el donatiu de Twitter es feia en mode irrevocable i no exclusiva llicència a la Library of Congress, amb els drets del donant de transferir o llicenciar sota els termes de Twitter tant en el moment de la donació com abans, en la llavors com la efectiva data. I els termes del servei de Twitter hi anaven annexes també. I que en quan a l'accés es donava un marge de sis mesos fins que Twitter cedís el permís a la Library of Congress d'exposar els tuits, en qualsevol format electrònic, sempre amb un ús raonable dels fitxers *robots.txt*, ja que Twitter no en permet en aquest contracte que la Library of Congress faciliti la descàrrega massiva de tuits. El tema de la disponibilitat, es va acordar que malgrat que la Library of Congress no cregués que ha de disposar d'alguna part de la col·lecció se li facilitarà de totes maneres. Ja que en l'apartat de garanties s'assegurava que la possessió i propietat física de la col·lecció serà lliure i clara de tota connexió, i el donant (és a dir Twitter) tindrà tot els drets, el poder i l'autoritat de transferir físicament la propietat de la col·lecció i els drets de llicència a la Library tal i com es descriu en el contracte. Per acabar, el contracte finalment se sotmet a les lleis dels Estats Units, i sota la legitimació de la cort federal del Districte de Colúmbia.

El 28 d'abril d'aquell mateix any s'aclarí una mica més l'anunci amb una entrada al blog oficial escrita per Matt Raymond director de comunicacions de la Library of Congress del 2006 al 2011. En aquesta entrada s'hi explicava com s'havien estat fent proves amb projectes com Twitterfall, un sistema web per veure tendències, cerques personalitzades, i d'altres actualitzacions en quasi temps real. Però sobretot s'intentava donar algunes dades més a l'anunci fet feia unes setmanes, com per exemple s'argumentava la importància d'haver triat Twitter, triant com a data rellevant per exemple la de l'elecció presidencial del 2008 amb Obama (com nosaltres també hem fet en l'apartat de "Diplomàtica amb Twitter"), o com els diaris incloïen en les seves cròniques tuits de Twitter com a font. S'explicava que l'acceptació de la donació de Twitter es feia per l'interès de la Library of Congress de preservar el coneixement i el seu context públic, i que per molt insignificants que fossin els tuits com a unitat documental, vistos en agregació formaven un recurs per a que les noves generacions entenguessin la vida en el segle XXI. A més s'aclaria que la Library of Congress no pagava per aquest arxiu «The Library did not pay for the archive; rather, it was a gift from Twitter» (la Library no va pagar per aquest arxiu; va ser més aviat un regal de

125 Dorsey, Jack. "just setting up my twtr." Twitter. <https://twitter.com/jack/status/20> (consultat a 2 de maig de 2014).

126 GIFT AGREEMENT <http://blogs.loc.gov/loc/files/2010/04/LOC-Twitter.pdf> (consultat a 25 d'abril de 2014)

Twitter). I que Twitter donava només els tuits públics, no pas els eliminats ni tampoc dades privades dels comptes de Twitter, tampoc la informació enllaçada ja fossin imatges o pàgines web, i la Library of Congress tampoc tenia planejat arribar a aquest nivell de profunditat. No s'especificava res de les fotografies adjuntades perquè com ja hem dit anteriorment aquesta característica audiovisual no s'implantà a Twitter fins al 2012. En l'entrada del blog s'acabava dient que l'objectiu primer de la Library of Congress era implantar el sistema que preservés a llarg termini aquesta col·lecció de tuits, i que seria després quan es consultaria amb experts i investigadors la manera d'exposar-la i oferir-la a consulta, protegint tant la privacitat com els drets d'explotació. No provarien de reproduir Twitter, sinó que la seva primera idea era la d'oferir els tuits de manera complementària a la d'altres col·leccions tals com la "National Elections Web Archive" o la "Supreme Court Nominations Web Archive".¹²⁷ Altres agències com la NASA ja alertaven sobre aquesta problemàtica de com presentar els tuits de Twitter i la interfície web: «NASA commented that Twitter feeds without the Twitter interface would likely read just the same as feeds within Twitter; as long as the content and essential properties are preserved, the value is retained»¹²⁸.

Malgrat tot seran uns 55 milions de missatges diaris, i no només serà dels personatges més populars, sinó que s'hi trobaran missatges de gent més corrent i menys elitista, cosa que fa que el projecte tingui enorme rellevància en el món acadèmic historiogràfic, ja que es podrà obtenir quasi el segon per segon de tot un registre històric de gent normal i corrent que voluntàriament està exposant la seva vida amb certa continuïtat, segons opinen gent com Fred R. Shapiro, bibliotecari de la Yale Law School¹²⁹. A més s'afegirà al ja de per si enorme espai de 167 terabytes de material digital de captures web, o els 21 milions de llibres de text de la biblioteca. I les implicacions però de tot plegat poden modificar el comportament dels usuaris de Twitter ja que com comentà Paul Saffo, professor a la School of Engineering de Stanford University, aquests podrien repensar-se les seves indiscrecions ara que passaran a formar part del registre històric i seran preservades durant generacions i generacions per a tot tipus d'investigadors.

Cap al 4 de gener de 2013 la Library of Congress anuncià que estaven assolint l'objectiu de rebre i preservar els bilions de tuits del 2006 al 2010¹³⁰, i que aquest seria el seu primer lot en el que treballarien fins a establir un sistema d'organització i de recepció fiable per a que assoleixin el seu objectiu de preservar diàriament tuits en l'actualitat. A més l'allau d'investigadors interessats en el projecte va pujar fins les 400 peticions, sobretot investigadors del periodisme ciutadà, de les

127 Raymond, Matt. "The Library and Twitter: An FAQ." Library of Congress Blog. <http://blogs.loc.gov/loc/2010/04/the-library-and-twitter-an-faq/> (consultat a 26 d'abril de 2014)

128 "A REPORT ON FEDERAL WEB 2.0 USE AND RECORD VALUE ." National Archives Records Management Information Page. <http://www.archives.gov/records-mgmt/resources/web2.0-use.pdf> (consultat a 3 de maig de 2014) pg. 16

129 Lohr, Steve. "It's History, So Be Careful Using Twitter." The New York Times. <http://www.nytimes.com/2010/04/15/technology/15twitter.html> (consultat a 26 d'abril de 2014)

130 Allen, Erin. "Update on the Twitter Archive at the Library of Congress." Library of Congress Blog. <http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/> (consultat a 2 de maig de 2014)

comunicacions de funcionaris electes (polítics devien referir-se), a més de temes mèdics i de predicció de l'activitat del mercat de valors.

A més publicaren un resum de tots els objectius que havien anat assolint al llarg d'aquell temps¹³¹. Informant de les dificultats tecnològiques d'oferir aquella col·lecció a un públic tant variat. Anunciaven també alguns termes que no apareixien en el contracte de cessió amb Twitter que van publicar al 2010, com era el de que el material addicional que vingués de Twitter estaria sota els termes que van acordar-se però que si les dues parts ho acordaven aquests podien anar sota uns termes diferents abans de rebre's aquests materials addicionals. S'informava que en l'acord del desembre de 2010, era la companyia Gnip (establerta a l'estat de Colorado) la que s'encarregava de moure les dades a la Library of Congress, ja que així ho havia nomenat Twitter. Aquesta transferència s'inicià al desembre de 2011 i cap al 28 de febrer de 2012 la Library ja havia rebut prop de 2,3 terabytes de dades comprimides, que un cop descomprimides ocupaven un total de 20 terabytes, contenint uns 21 bilions de tuits¹³², amb prop de 50 camps de meta-dades, tals com localitzacions i descripcions. Cap a l'1 de desembre de 2012, ja s'haurien rebut més de 150 bilions més, per un total en l'arxiu 2006-2010 de 170 bilions de tuits que ocupaven uns 133,2 terabytes en dues còpies comprimides.

Un cop es va assegurar aquesta rebuda va iniciar-se també l'establiment de l'adquisició del que es publicava contínuament a Twitter, per tal de tenir un flux continuat de continguts digitals que es transferissin de Twitter cap a la Library. Seria Gnip qui designada per Twitter rebria els tuits en temps real, els segmentaria per hores, i els pujaria en un servidor segur esperant la petició de transferència de la Library. Llavors la Library descarregaria els fitxers a un espai del servidor temporal, comprovaria que el material fos complet i no hi hagués cap transferència corrupta, capturaria estadístiques sobre el número de tuits de cada fitxer, copiaria el fitxer en cintes, i eliminaria el fitxer de l'espai temporal del servidor. Aquesta infraestructura tècnica segons la Library és la que se segueix amb d'altres col·leccions digitals. Les cintes són el mètode estàndard d'arxivament en preservació de llarga durada. Els fitxers són copiats en dues cintes que s'arxiven en diferents localitzacions geogràfiques, com a mesura de preservació i de seguretat.

La Library advertia a més que podien passar mesos, inclòs anys fins que una adquisició és inventariada, organitzada i catalogada amb la informació dels materials accessible per als investigadors. La novetat amb la col·lecció de Twitter és que el ritme de creixement d'aquesta és cada dia més ràpid (almenys de moment) i la varietat dels tuits va en augment sobretot ara que s'han de contemplar continguts audiovisuals que van adjunts als tuits. Però és que asseguren que per ara les recerques que han provat a nivell intern en el material rebut fins ara, els resultats

131 "Update: Twitter Archives, Library of Congress." Library of Congress.

http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf (consultat a 2 de maig de 2014)

132 Recordar que els bilions al món anglosaxó són desenes de milers de milions, però no en el món nord-americà.

tarden unes 24 hores en rebre's, i s'està treballant en computació paral·lela i distribuïda per tal de millorar aquesta marca, tot i que els serà difícil al tractar-se d'una infraestructura que suposaria centenars si no milers de servidors, un cost no assumible pràcticament per una institució pública. Ells mateixos diuen que moltes empreses privades estan prometent l'accés a aquest historial de tuits, però no és un accés lliure, indexat ni investigable de manera prou fiable per a ser vàlid i útil als investigadors i estudiants.

Cap al 25 d'octubre de 2013 (una setmana abans de Halloween), la NARA publicava el "NARA Bulletin 2014-02" que eren les directrius per a la gestió dels registres de mitjans de comunicació social, és a dir de com s'havien de gestionar els ingressos de la documentació rebuda de les xarxes socials. Aquest butlletí reemplaçava l'anterior "NARA Bulletin 2011-02: Guidance on Managing Records in Web 2.0/Social Media Platforms", i pel títol ja vam deduir que el terme "Web 2.0" que tant de moda s'havia posat no fa masses anys quedava ja descartat en la nova etapa.

En les directrius s'establien les bones pràctiques en la captura i registre dins de les agències federals de l'ús que en feien dels mitjans de comunicació social, la seva legislació, regulació i pòlisses, però no es tractava d'una guia sobre cap plataforma específica de captura de continguts. En la seva definició de mitjans de comunicació social s'inclouïen blogs, wikis, xarxes socials, arxius de fotografies, mons virtuals, serveis basats en localitzacions, i compartició de vídeos. I seguint amb la definició afegia que un contingut social era aquell que permetia col·laborar, crear, organitzar, editar, comentar-lo, combinar-lo, compartir-lo, tal i com establia la The Federal Records Act (44 U.S.C. 3301), que designava tot aquell material enregistrat, que fet o rebut en el decurs de les activitats de les agències federals, més enllà de les seves formes característiques, havia de ser preservat. D'una manera més simple establia quatre preguntes per definir quins continguts havien d'acabar preservant-se:

- Conté evidències de les polítiques, gestions i missió de les agències?
- Està disponible la informació només en un lloc de mitjans de comunicació social?
- Utilitza l'agència alguna eina per transmetre informació oficial de l'agència?
- Hi ha alguna necessitat econòmica per la informació?

Si qualsevol d'aquestes preguntes es responia afirmativament, allò podia arribar a esdevenir amb certa seguretat un registre federal. Però no ho acabaria sent fins que no fos investigable, es pogués afegir-hi comentaris i estigués disponible per a les col·laboracions, que tingués el seu contingut definit, el seu context, la seva estructura associada a metadades (autor, data de creació,...), i que permetés la seva fiabilitat i autenticitat.

El butlletí però no aclaria què passava amb les col·leccions de mitjans de comunicació social,

quan aquests mitjans seguissin actius diàriament, per exemple què passaria si un tuit emmagatzemat per la Library of Congress seguís rebent respostes i contingut addicional com es faria per oferir la versió més actualitzada d'aquell tuit. A més especificava que alguns missatges no s'havien de guardar, com els considerats SPAM, cosa que donaria més pulcritud als registres però distorsionaria la realitat del que suposaven els continguts d'aquestes xarxes socials. Tot i que ja avisava que les agències s'haurien de fer càrrec de completar els requeriments de les captures en cas que les companyies de tercers encarregades se'n desentenguin. Però les complicacions de què fer amb una col·lecció que s'anava modificant diàriament eren reconegudes també per la pròpia NARA:

«As the complexity of social media becomes greater, the ability to capture and preserve content will also become more complex. Some tools lend themselves more easily to a complete harvest of information; Twitter threads, for example, can be removed from the Twitter site and preserved in another storage location.»¹³³

Entre les eines per capturar aquestes col·leccions, s'esmentaven rastrejadors web per crear versions locals dels llocs web, plataformes que feien ús de les interfícies de programació de les aplicacions (APIs), també l'ús dels fils RSS i exportació de continguts.

La NARA en el seu "White Paper on Best Practices for the Capture of Social Media Records"¹³⁴ del maig de 2013 parlava també de la preservació de Twitter. I dins de l'apartat "IV. Examples of social media capture tools and methods" enumerava to d'eines per explotar els mitjans d'informació socials, tals com Hootsuite o IFTTT. Però també eines de còpia, d'arxivament, indexació, seriació com Gloal Relay¹³⁵ o Hanzo Archives¹³⁶, Recollect¹³⁷ i fins i tot Patrina¹³⁸ que és un servei estil WORM¹³⁹. Eines que fins i tot permeten programar captures d'aquesta informació, alhora que afegir-hi marques d'aigua i signatures a les captures de pantalla. O com el servei Smarsh¹⁴⁰ que sota subscripció permet capturar, preservar i indexar les publicacions que fem a Twitter, a través de la API en temps real, amb segell temporal, hash i còpia encriptada en DVD,

133 "A REPORT ON FEDERAL WEB 2.0 USE AND RECORD VALUE ." National Archives Records Management Information Page. <http://www.archives.gov/records-mgmt/resources/web2.0-use.pdf> (consultat a 3 de maig de 2014) pg. 19

134 Davis-Castro's, Carla. "Best Practices for Social Media Capture." National Archives Records Management Information Page. <http://www.archives.gov/records-mgmt/resources/socialmediacapture.pdf> (consultat a 3 de maig de 2014)

135 "Messaging & Archiving Solutions for Compliance | Global Relay." Messaging & Archiving Solutions for Compliance | Global Relay. <http://www.globalrelay.com> (consultat a 18 de maig de 2014)

136 "Defensible Web Archiving - Hanzo Archives." Hanzo Archives. <http://www.hanzoarchives.com> (consultat a 18 de maig de 2014)

137 "Recollect." - Beautifully Archive and Explore Your Digital Life. <http://recollect.com> (consultat a 18 de maig de 2014)

138 "Social Media Archiving." Patrina. <http://www.patrina.com/data-management-archiving/social-media-archiving/> (consultat a 18 de maig de 2014)

139 *Write Once Read Many*. Que s'escriu una vegada però es pot llegir múltiples vegades. És a dir només escriptura i lectura, però no modificació del contingut.

140 "Archiving Solution | Archiving & Compliance | Smarsh." Smarsh RSS2. <http://www.smarsh.com> (consultat a 18 de maig de 2014)

servei que també ofereix SMC4¹⁴¹. D'altres com SocialSafe permeten guardar per exemple tots els tuits en un fitxer comprimit ZIP amb les fotografies adjuntes als tuits i un fitxer CSV amb els tuits, el que ja ens ofereix gratuïtament Twitter però SocialSafe ens ho ofereix en versió de 30 dies de prova o un preu mínim de 5,99\$¹⁴² l'any. La llista seguia amb exemples de codi obert com ThinkUp¹⁴³ o d'altres com Total Discovery¹⁴⁴ que aposten per fins i tot oferir suport legal en cas que volem usar els missatges dels mitjans socials en algun litigi. No és fins a la part final d'aquest breu manual de la NARA que ens apareixien ja algunes eines per a dur a terme l'arxivament dels nostres missatges a Twitter usant els nostres propis mitjans i les eines que ja ens ofereix pròpiament Twitter tals com la API o la petició de descàrrega d'un "arxiu" amb els nostres missatges.

Les recomanacions finals de la NARA però eren les d'utilitzar eines el més transparents possible i sobretot respectar les APIs que ens ofereixen els serveis com Twitter per tal de capturar adequadament els continguts que se'ns ofereixen, ja que hem d'evitar topar-nos amb els termes legals de les xarxes socials com Twitter que limiten certs usos del que n'obtenim a través de la API o d'altres sistemes.

3.2 L'exemple dels National Archives del Regne Unit

Els National Archives del Regne Unit també es van apuntar a catalogar els continguts que els organismes governamentals publiquen a les xarxes socials¹⁴⁵. També amb l'argument de registrar i tenir captures dels moments més destacats de la història anglesa, calculant que de moment tenien en ment capturar uns 65.000 tuits del govern central i també dels organismes que van dirigir els últims Jocs Olímpics i Paralímpics a Londres el 2012 (per tant registren des del 2008 al setembre de 2013), que serien preservats permanentment en registres públics.

La diferència i el que fa especial el projecte dels National Archives és que ofereixen ja de cop vàries opcions de visualització d'aquests tuits¹⁴⁶. Una visualització en JSON, una en XML i l'altra és un enllaç directe al compte de Twitter. Aspecte el de la visualització que fins ara cap projecte ens havia mostrat, ja que o ho estaven discutint o desenvolupant.

Aquest projecte dels National Archives sembla que pretén tenir certa continuïtat i que s'aniran

141 "SMC4 Social Media Capture, Control, Communication and Compliance." SMC4 Social Media Capture, Control, Communication and Compliance. <http://www.smcapture.com> (consultat a 18 de maig de 2014)

142 "Completely free for PC & Mac." Pricing. <http://www.socialsafe.net/pricing> (consultat a 3 de maig de 2014).

143 "ThinkUp." ThinkUp. <https://www.thinkup.com> (consultat a 18 de maig de 2014)

144 "eDiscovery, Electronic Discovery, Computer Forensics, Data Discovery, Legal Holds." BIA. <http://www.biaprotect.com> (consultat a 18 de maig de 2014)

145 Williams, Rhiannon. "Tweets and social media to enter National Archives." The Telegraph. <http://www.telegraph.co.uk/technology/news/10813826/Tweets-and-social-media-to-enter-National-Archives.html> (consultat a 2 de juny de 2014)

146 "UK Government Web Archive: Twitter." The National Archives. <https://www.nationalarchives.gov.uk/webarchive/twitter.htm> (consultat a 2 de juny de 2014)

monitoritzant tots els tuits produïts pel govern del Regne Unit. Sempre tenint present la preservació digital, l'accessibilitat, la retenció del context i la disposició per a la reutilització.

En qüestió interna els tuits publicats pel govern central i les organitzacions oficials dels Jocs Olímpics i Paralímpics de 2012 seran capturats i els enllaços que continguin seran inclosos també al UK Government Web Archive per tal que els usuaris consultats puguin consultar també aquestes pàgines web enllaçades independentment de si continuen "en línia" o no.

A nivell extern hem vist que s'exclouen els retuits i tuits enviats des de comptes no relacionats amb el govern i que per tant no formen part de cap conversa del govern i no apareixen. Tampoc s'arxivaran les webs externes que ja estiguin en d'altres sistemes d'arxivament com les de pàgines de notícies, i en aquest cas el que es farà és incloure un missatge 410 (error HTTP 410 o sigui que el recurs telemàtic ja no es troba disponible i no tornarà a estar-ho¹⁴⁷) però que a més indicarà la destinació dels enllaços per si algú vol localitzar els continguts on sigui.

Els continguts recopilats pels National Archives tal i com indiquen a les clàusules de "Re-use of content accessible through the UK Government Web Archive"¹⁴⁸, no podran ser reutilitzats tant lliurement com els generats pels propis organismes governamentals, ja que els tuits de Twitter estan sota els termes d'ús de Twitter i serà per aquests pels que ens haurem de regir.

3.3 L'exemple de la Facultat de Biblioteconomia i Documentació

Al 2012 es va publicar un article titulat "Seguiu-nos: estratègies de captura de tuits en català"¹⁴⁹ per Anita E. Locher, Aina Giones-Valls, Eli Ramírez i Greda Ortiz, on s'analitzaven també les possibilitats que podia oferir Twitter per preservar-ne els missatges des d'una perspectiva de gestió biblioteco-documental, proposant una estratègia de captura de tuits per a la Biblioteca de Catalunya, que incloïa criteris de selecció, metadades de preservació i previsions de creixement. L'estudi es basava també com nosaltres hem fet en la informació aportada per la Library of Congress, però ells també varen consultar Twitter'n'Català i el PADICAT (Patrimoni Digital de Catalunya). L'article elogiava la difusió i ús tant estès de Twitter a Catalunya, i també com hi havia diversos projectes que estaven capturant contingut de Twitter.

Per altra banda la seva preocupació al preservar els tuits s'originava perquè segons ells Twitter només permetia accedir als últims 3200 tuits d'un usuari i als 1500 tuits per cada cerca. Això ja no és així, i com ja hem comentat ara cada usuari pot descarregar-se tot l'arxiu complet dels seus

¹⁴⁷És usat per indicar que un contingut ha estat purgat intencionadament. Wikipedia contributors, "List of HTTP status codes," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=List_of_HTTP_status_codes&oldid=609124769 (consultat a 2 de juny de 2014)

¹⁴⁸ "Information on web archiving." The National Archives. <https://www.nationalarchives.gov.uk/webarchive/information.htm> (consultat a 2 de juny de 2014)

¹⁴⁹ E. Locher, Anita, Aina Giones-Valls, Eli Ramírez, and Greda Ortiz. "Seguiu-nos: estratègies de captura de tuits en català." BiD: textos universitaris de biblioteconomia i documentació, desembre, núm. 29. <http://bid.ub.edu/29/locher1.htm> (consultat a 3 de maig de 2014)

tuits, per tant la premissa ja no és vàlida actualment, però podem seguir analitzant la resta de l'article.

En la seva metodologia es van dedicar a preguntar a membres de la International Internet Preservation Consortium quina metodologia feien servir per la preservació de la informació de les xarxes socials. En el nostre cas no vam haver de rondar, la Library of Congress s'emporta la col·lecció així que varen ser ells a qui vam consultar, sense massa resposta més enllà d'un breu correu electrònic que ens remetia a consultar el que exposaven al seu blog i d'altres webs seves que és segons ells on exposarien tota la informació disponible destinada a ser difosa públicament.

Tot seguit l'article exposava com per al filtre i tria del que es preservava s'optava per criteris de país de producció, ciutadania, idioma o temàtica. Les seves conclusions era que s'havia d'usar diverses eines per tal de filtrar, capturar i preservar els tuits. Es comentava a l'article que no es coneixia cap institució que estigués acumulant dades i informació de Twitter. Bé, al 2012 es desconeixia segurament la campanya per exemple de Cesticat¹⁵⁰ o la de politwoops¹⁵¹, ambdues fetes públiques el 2013.

En la seva tria d'eines exposaven sobretot les que permetien visualitzar els perfils de personatges més populars com TwitterGrader, o les que permetien filtrar resultats segons l'idioma com Twitter 'n' Català. També es parlava de l'ús de l'API de Twitter per obtenir la informació i les meta-dades, però es preferia eines d'arxivament com TwapperKeeper o Backupify, les quals tindrien poca competitivitat amb el #TweetArchive que hem explicat anteriorment que és el servei de Twitter per rebre el nostre "arxiu" de tuits. A més algunes d'elles com la Heritrix¹⁵² no tenen la documentació sobre Twitter actualitzada des del 2012 per tant dubtem que pugui tenir encara cap ús amb la versió 1.1 de la API de Twitter¹⁵³ i més quan el que volem és les dades en el seu format original no pas la versió en HTML, cosa que Twitter ja ens ofereix fins i tot en CSV i JSON.

Com exposa l'article, Heritrix i per tant Padicat només fan una còpia del que es visualitza al navegador, un rastre inert del que s'havia pogut veure en el seu moment de la web, però li manquen les entranyes, el que ocupava i es mantenia junt formant el cos del missatge o document a mostrar. Tot seguit s'exposaven les dificultats per capturar cerques (basades en *hashtags*) quan

150 "Mitjans i periodistes en les llistes de seguiment virtual dels Mossos" Media.cat - Observatori crític dels mitjans. <http://www.media.cat/2013/10/28/mitjans-i-periodistes-en-les-llistes-de-seguiment-virtual-dels-mossos/> (consultat a 3 de maig de 2014)

151 "Politwoops - All deleted tweets from politicians." politwoops. <http://www.politwoops.com/g/Parlamentaris%20catalans> (consultat a 4 de maig de 2014) anunciada el 17 de desembre de 2013: "Politwoops, el caçador de piulades esborrades de polítics." VilaWeb.cat. <http://www.vilaweb.cat/noticia/4162478/20131217/politwoops-cacador-piulades-esborrades-politics.html> (consultada a 11 de juny de 2014)

152 Levitt, Noah, and Adam Miller. "Facebook and Twitter Scroll-down" Heritrix - IA Webteam Confluence. <https://webarchive.jira.com/wiki/display/Heritrix/Facebook+and+Twitter+Scroll-down> (consultat a 3 de maig de 2014)

153 Sippey, Michael. "Changes coming in Version 1.1 of the Twitter API | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2012/changes-coming-to-twitter-api> (consultat a 3 de maig de 2014)

es feia amb el Heritrix¹⁵⁴, per després lamentar que l'ús de *robots.txt* limitava la captura de pàgines en HTML de Twitter. S'argumenta que les limitacions de la API de Twitter són les que fan que Padicat no pugui capturar les dades, cosa que és certa, que hi ha limitacions però la consulta i treball amb les dades és possible i fructífera, altra cosa és que sigui quantificable econòmicament. I fins i tot, afegeixen a l'article una justificació en el seu interès per capturar Twitter dient que «la Llei 4/1993, de 18 de març, del sistema bibliotecari de Catalunya permet recopilar, conservar i difondre la producció bibliogràfica catalana incloent-hi la digital». També exposen els problemes ètics de Twitter, tot i que més aviat seria el tema de la credibilitat del que ja hem parlat en l'apartat concret (*La credibilitat de Twitter*) dedicat a aquest debat.

Com a treball més pràctic van provar de capturar en HTML cerques per localització (*place*) i idioma (*language*). Com ja hem dit no seria un plantejament vàlid quan Twitter ofereix les dades en JSON, XML i CSV. Però van acabar resolent que era millor contactar amb la Library of Congress i que l'IIPC desenvolupés més projectes comuns per facilitar aquestes dades a l'entorn d'usuaris catalans. Com a formats però s'apostava per PDF («és un dels formats considerats sostenibles que manté els enllaços actius»), JSON i HTML (tot i que JSON sigui un llenguatge de marcatge i no un format).

La solució final que proposaren fou la de delimitar els Tuits públics de l'Administració, Entitats catalanes, Portals, Mitjans de comunicació d'àmbit geogràfic català, Museus catalans, Biblioteques i arxius, Partits polítics catalans, i a més crear un compte específic de preservació que guardés automàticament els tuits dels seus seguidors. Es continuava apostant per Heritrix del qual se n'esperava una versió 1.14.4 que solucionaria tots els problemes esmentats més amunt¹⁵⁵. Tot seguit es continuava teoritzant sobre la freqüència de captura¹⁵⁶, i es feia una aproximació de la mida d'un tuit en 1,5Kb¹⁵⁷, per acabar pronosticant encara més requeriments d'espai quan s'adjuntessin vídeos als tuits cosa que actualment amb el servei Vine ja succeeix però no sabem encara com Twitter ho implementarà en les metadades, ja que de moment inclou una direcció web a l'esquema de metadades i prou.

3.4 Quina podria ser la meva proposta

Tot seguit, un cop hem avaluat els casos i coneixent les capacitats que ens ofereix Twitter, podríem teoritzar sobre quines possibilitats trobaríem nosaltres aprofitables en el terreny de

154 Als Annexos quan parlem de com treballar amb la API a "Manual per extreure dades en JSON amb Postman i l'API de Twitter" s'ofereix informació per fer-ho adequadament sense passar per Heritrix.

155 Bàsicament se'm fa difícil d'entendre com un rastrejador de pàgines web com Heritrix seria capaç d'accedir a la base de dades de Twitter i extreure'n totes les dades saltant-se les mesures de seguretat. Però potser se'm passa per alt alguna capacitat d'aquest programari que desconec.

156 Un tema poc rellevant quan a través de la API Twitter t'ofereix tota la informació quasi al moment que la demanes.

157 No sé d'on treuen aquesta mida quan a mi em surt a més del doble d'aquesta mida en JSON i una mica més d'aquesta mida fins i tot quan rebem les metadades del tuit en text en cru.

l'arxivística amb Twitter.

Com hem explicat a la introducció del treball cal que tinguem molt clar què és Twitter, i com es formen els documents que anomenem tuits. Si no tenim un coneixement profund d'això no ens podem posar a especular sobre el que seria la nostra proposta.

Segons la meua visió personal, com he exposat al llarg d'aquest treball sobre Twitter s'ha de fer des del codi, des de les entranyes del document i després ja passarem a contextualitzar-lo per tal de completar el tractament arxivístic amb tots els requeriments que exigíssim. És un error creure que el que ens mostra projectat el navegador ha de ser forçosament tot el document, a no ser que desitgem viure en la imatge interessada que se'ns ofereix¹⁵⁸. Almenys en la meua proposta no em conformo amb la captura de pantalla que un navegador m'ofereix, espero l'accés al document complet i aquest es troba a la base de dades MySQL que gestionada pel PHP ens arriba en HTML.

Hem de tenir molt clar que quan vam acceptar Twitter com a "eina" de treball vam acceptar-ne també les seves condicions, els seus termes d'ús, i una d'aquestes era no apropiarnos de la base de dades de Twitter de manera massiva sense l'autorització de Twitter, cosa que molts projectes de treballar arxivísticament amb Twitter sembla que pretenguin. Per tant, estem molt d'acord amb com es va iniciar el projecte entre la Library of Congress i Twitter, amb un mutu acord de gestió autoritzada de la base de dades, no pas amb l'extracció unilateral i indiscriminada via API.

Si fem memòria, al cap de poc d'iniciar-se Twitter, en el que es concep com la primera època¹⁵⁹, varen sorgir alternatives que trencaven aquesta hegemonia de Twitter, i que ens permetien posseir tant una eina de similars característiques com rebre'n la possessió dels drets d'explotació. Es tractava de serveis de programari lliure com Identi.ca (que passaria a anomenar-se Laconica i ara pimp.io) o Jisko (creat per Rubén Díaz i Gabriel Longás al 2008¹⁶⁰ i que després d'un temps de finançament incert¹⁶¹ va acabar tancant al 10 de setembre de 2010¹⁶²). Malgrat patir

158 Que no es passi per alt que sí, he colat certa referència al Mite de la Caverna de Plató.

159 Es tenen més o menys distingides varies èpoques: Del 2006 al 2008 és l'època dels desenvolupadors i persones molt interessades en la informàtica, del 2008 al 2009 és l'època dels periodistes més generalistes, i a partir del 2009 és l'època de la gent famosa o de vida més pública. Finalment també es compta el 2010-2011 com l'època en que s'omple d'usuaris amb un nivell més baix de coneixements generals per dir-ho d'alguna manera, i el 2012 com l'eclosió dels que comparteixen tot tipus de continguts multimèdia i que Twitter ja no és només els 140 caràcters (més les metadades). Aclarir que això no categoritza definitivament cada usuari, però dona una idea de les onades que ha anat rebent Twitter.

160 Ramón Mora, Juan. "Jisko abierto a todos (La fábula de la ardilla y la ballena) | JRMora, humor gráfico." JRMora humor gráfico RSS. <http://www.jrmora.com/blog/2008/08/01/jisko-abierto-a-todos-la-fabula-de-la-ardilla-y-la-ballena/> (consultat a 4 de maig de 2014)

161 Yuste Robles, Bárbara. "ABC.es." ABC. <http://www.abc.es/20090701/medios-redes-web/jisko-twitter-espanol-pide-200907011942.html> (consultat a 4 de maig de 2014).

162 Bortnik, Sebastian. "Jisko: crónica de una muerte anunciada - Bitelia." Bitelia. <http://bitelia.com/2010/09/jisko-cronica-de-una-muerte-anunciada> (consultat a 4 de maig de 2014).

les dificultats de no tenir un suport econòmic massa gran ni rebre massa atenció pública, van anar reactivant-se i desactivant-se, transformant-se i encara podríem trobar-ne versions i nous projectes creats a partir del codi d'aquests serveis de programari lliure. Aquestes alternatives venien a respondre la inquietud per la "possessió dels tuits", és a dir, qui guanyava la pugna entre el que compartim a Twitter i el que Twitter pot beneficiar-se a partir del que nosaltres hi publiquem.

Sonarà estrany que sortim amb aquest tema quan hem anomenat varies vegades projectes que no mostraven cap reticència a xuclar i capturar a discreció tot el que els plagués de Twitter, però els termes de l'API de Twitter són ben clars¹⁶³: A la secció "I. Twitter Content" al punt "4. A" ens informa que «Exporting Twitter Content to a datastore as a service or other cloud based service, however, is not permitted.», per tant no podem emmagatzemar en un espai accessible a Internet el que exportem de Twitter a través de la API. També al punt "6." ens informa que «You do not have a license to Twitter Content submitted through your Service other than the rights granted in the Rules.», per tant no tenim cap llicència ni dret garantit per publicar les dades de Twitter a través d'un servei web que creéssim a Internet.¹⁶⁴

Al punt "4. Be a good partner to Twitter" se'ns diu que has d'indicar que els tuits que has extret i que exposes provenen de Twitter, indicant de quin usuari i aportant el material gràfic que ho acrediti (logotips, avatars, sobrenoms,...), has de respectar l'estructura de Twitter i no desactivar-ne cap funcionalitat al mostrar els tuits si ho fas incrustant-los fora de Twitter (no es permet qualsevol captura), s'ha de respectar la propietat intel·lectual i els drets dels usuaris, i no s'ha d'aportar cap element distintiu que no aparegui a través de la API de Twitter. I a "IV. Commercial Use" s'especifica com no es pot fer cap ús comercial dels tuits que s'extreuen fora de Twitter. A més, a l'apartat "3. Ownership; Feedback." ens recorda com ja se'ns havia dit en els termes del servei, que si treballem amb Twitter estem al cas i tenim tot el coneixement que ells retenen tots els drets mundialment dels continguts de Twitter, incloent-ne la propietat intel·lectual. No es pot qüestionar aquesta validesa de la possessió dels drets (sic) i no es pot desafiar l'explotació dels continguts sense autorització. A més tu cedeixes a Twitter els drets de la còpia, modificació, creació derivada, difusió de publicitat, revelació, distribució, llicència i sub-llicència, incorporació i qualsevol altre mena d'interacció, incloent obres derivades per vies comercials o no comercials i propòsits sense cap obligació per part teva. Finalment a l'apartat "4. Updates.", Twitter es reserva la capacitat de modificar les normes d'ús de la API, els termes i condicions, incloent els requeriments de com es poden mostrar els tuits, i que si no estàs d'acord amb els canvis seràs lliure de cessar el teu ús de la API i dels continguts de Twitter, ja que si continues usant l'API de Twitter i els seus continguts estaràs acceptant de facto els canvis.

163 "Developer Rules of the Road." Twitter Developers. <https://dev.twitter.com/terms/api-terms> (consultat a 4 de maig de 2014).

164 "API Terms of Service: Diff between Sept 5, 2012 and July 2, 2013." Twitter Developers. <https://dev.twitter.com/terms/api-terms/diff-20130702> (consultat a 28 de maig de 2014)

Per tot això ens sorprèn aquesta perseverança en xuclar de Twitter, quan el propi Twitter és prou obert per a que no ens haguem de fer còpies locals de tot i duplicar informació que mentre existeixi Twitter tindrem accessible. I si algú en dubta pot esperar-se a que la Library of Congress acabi el seu projecte de col·lecció.

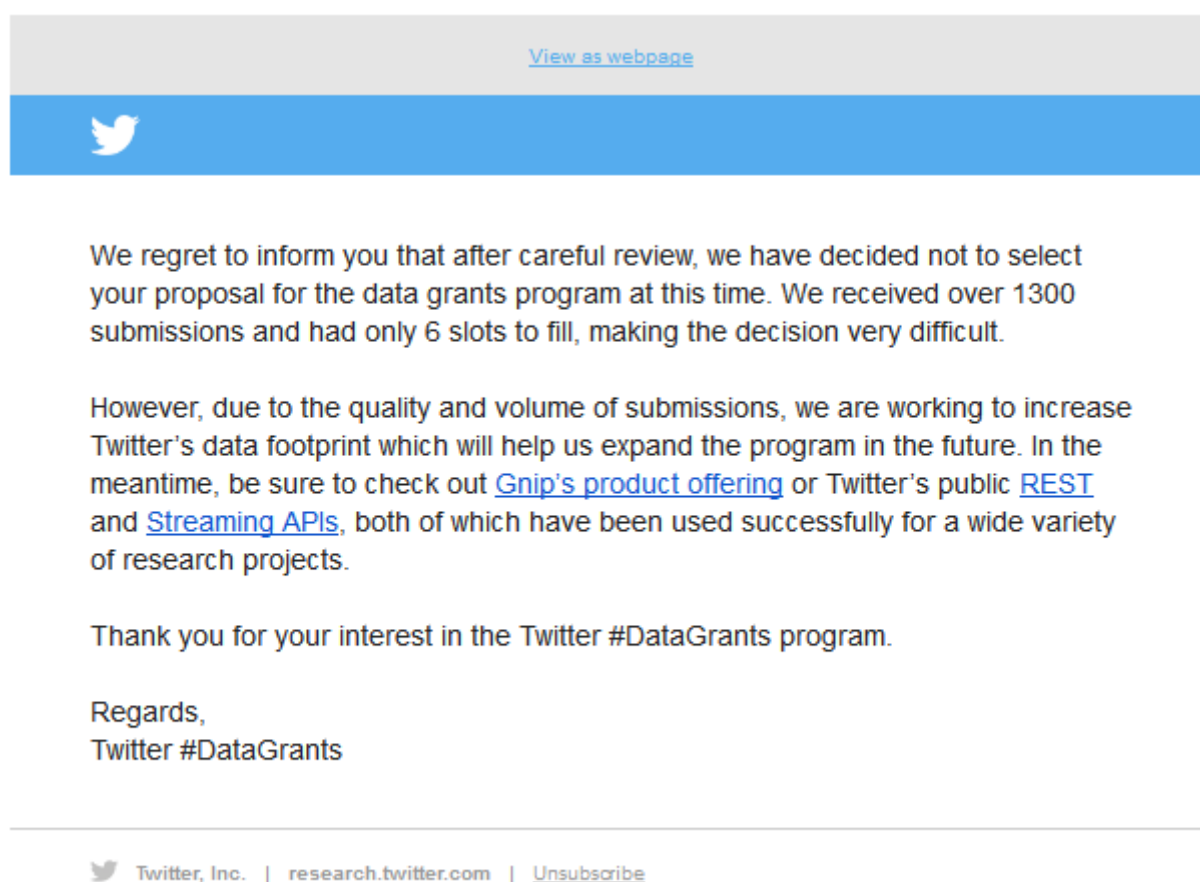
En aquest apartat del treball podríem parlar de les idees de Evgeny Morozov sobre assolir una espècie de sobirania popular de les xarxes socials, i a nacionalitzar aquest "Internet de les coses". Mostrar-nos contraris a «maximitzar l'eficiència per damunt de valors com la convivència o la llibertat dels individus»¹⁶⁵. Que ja que aportem dades i continguts a les xarxes socials, obtenim d'elles la possessió i el control d'allò que hi hem publicat. De posar les infraestructures dels *data centers* i grans centres de processament sota control dels ciutadans, bàsicament Morozov planteja nacionalitzar les xarxes socials.

Plantejaments com el seu topen amb dues coses, però bàsicament amb una de sola. El que ell planteja és que els ciutadans es veuen empesos a compartir i per tant han de rebre una compensació que seria la seva sobirania sobre el que comparteixen, per l'altra planteja que es nacionalitzi els espais on es comparteixen les dades i continguts dels ciutadans. Això no s'entendria si som conscients que els ciutadans trien què comparteixen i què no. Ni Facebook ni Twitter requereixen de cap certificació sobre si les dades d'identificació personal que aportem són verificables, sinó que ens confia la nostra bona fe en saber identificar-nos fins a on creguem necessari. A més, sabem que no tothom té Facebook ni tothom té Twitter, per tant podem triar què compartim i triem on ho compartim. Els plantejaments de Morozov tot i que en la temàtica de *big data* i *smart cities* poden semblar interessants (en la seva vessant de *question men*), no ho semblen tant en el tema de des-privatitzar segons quins serveis que no estan tant socialitzats com ell pensa.

Dins aquesta voràgine de qüestionament de la possessió de les dades generades a les xarxes socials Twitter va fer també un pas endavant anunciant el seu programa Twitter Data Grants¹⁶⁶. Que pretenia ser un fons de dades públiques i històriques de l'empresa Twitter per aquelles organitzacions acadèmiques o centres d'investigació que ho sol·licitessin abans del 15 de març de 2014. Vam provar d'optar a aquest programa, però se'ns va denegar la petició per part de Twitter tot convidant-nos a esperar més places lliures en el futur.

165 Puigtobella, Bernat. "Evgeny Morozov: "Potser és l'hora de desprivatitzar Facebook i Google"." Núvol. <http://www.nuvol.com/noticies/evgeny-morozov-potser-es-lhora-de-nacionalitzar-facebook-i-google/> (consultat a 4 de maig de 2014)

166 Polo, Rebeca. "Twitter abre sus datos a académicos e investigadores." *wwwwhat's new*. <http://wwwwhatsnew.com/2014/02/07/twitter-abre-sus-datos-a-academicos-e-investigadores/> (consultat a 4 de maig de 2014).



3. Il·lustració: Missatge del correu electrònic que vam rebre de resposta el 17 d'abril de 2014 per part de Twitter Data Grants.

El que oferia Twitter era l'accés a paquets de dades segons el que declaréssim estar investigant. A més, s'oferia contacte amb els enginyers i investigadors propis de Twitter, i un espai on publicar-hi els nostres estudis.

Tot el sistema responia a l'acord entre Twitter i Gnip¹⁶⁷ que hem anomenat en l'apartat on parlàvem de la col·lecció de Twitter que rep la Library of Congress i que és la companyia Gnip la que en gestiona el trasllat. Desconeixem més dades ja que com hem dit no se'ns va acceptar en el programa, però suposem que es devien rebre "arxius" com el que s'ofereix a cada usuari en el seu perfil privat, però en el cas de Twitter Data Grants amb moltes més dades i de diversos usuaris, segons s'indiqués a la petició.

Continuant en la meua valoració d'estudis, el de Miguel Ángel Domingo Arnal "Modelado, extracción y análisis de información del flujo de datos de Twitter"¹⁶⁸ és potser dels més encertats i

167 Krikorian, Raffi. "Introducing Twitter Data Grants | Twitter Blogs." Twitter Blogs.

<https://blog.twitter.com/2014/introducing-twitter-data-grants> (consultat a 4 de maig de 2014)

168 Domingo Arnal, Miguel Ángel, and Julià Minguillon. "Modelado, extracción y análisis de información del flujo de datos de Twitter." Repositori institucional - UOC.

<http://openaccess.uoc.edu/webapps/o2/handle/10609/18961>

[<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/18961/8/mdomingoarnTFM0113memoria.pdf>] (consultat a 11 de maig de 2014)

pròxims a el que em sembla més idoni de treballar amb Twitter dels que he pogut llegir. Començant pel seu convenciment d'usar eines obertes per tal de ser el més auditable possible. Passant per la captura mitjançant l'API de Twitter. I com la informació acaba estructurada i útil per obtenir-ne conclusions.

A més, té una concepció de com funciona Twitter molt clara, i sap que les dades a Twitter provenen originàriament d'una base de dades, no pas de pàgines HTML. Per tant ens mostra com hem de treballar per tal d'interactuar amb Twitter i que ens doni el que volem saber d'aquesta base de dades (dintre dels límits de la API de Twitter lògicament). Diferencia entre la captura de tuits contínua per "stream" (flux continuu) o per "search" (cerca puntual). La captura per "Stream" permet seguir un esdeveniment actual però no permet obtenir publicacions passades, però a canvi s'obté moltes dades en forma de metadades (les entitats que hem anomenat a l'anàlisi amb Diplomàtica de Twitter) que pel sistema de "Search" no obtindríem.

Miguel Ángel parla de l'eina Phirehouse¹⁶⁹ desenvolupada en PHP, la qual permet anar descarregant en una base de dades els tuits que anem capturant de Twitter. L'autor obvia que a part d'un sistema de base de dades (en el seu cas tria MySQL com té també el propi sistema de Twitter) s'hauria de tenir instal·lat i configurar el sistema que interpretés el codi PHP, cosa que dificultaria l'ús de l'eina a gent amb pocs coneixements de programació. A més com alerta l'article es poden produir errors en la captura i haver de fer correccions en la base de dades degut a captures duplicades. Es veu que tampoc no s'acaba obtenint una base de dades tan clara com la que podem aconseguir amb la funció de descarregar un "arxiu" dels nostres tuits en CSV que hem comentat anteriorment, ja que amb Stream les dades que acaben a la base de dades no es distribueixen ben bé els camps ja que el sistema se sobrecarrega al fer-se tot "en viu", per tant el sistema no és tant flors i violes com es podria esperar. Però el sistema creat segons el que indica aquest article permet fins i tot actualitzar la informació dels perfils dels usuaris, una funció interessant ja que completaria tota la informació que s'obté en les consultes. A més, es munta un sistema de consultes a Google per determinar les coordenades de geolocalització (si és que apareixen en els tuits) i poder descriure amb topònims des d'on s'estan enviant els tuits.

Molt interessant el treball que en fa de les dades obtingudes i exposades en taules, ja que aconsegueix idear diversos indicadors, com l'Índex de Ressonància o quants retuits i respostes provoquen els tuits d'un usuari, o l'Índex de Creativitat o quants dels tuits publicats per l'usuari són originals de l'usuari¹⁷⁰. Aquestes aportacions poden ajudar molt a futurs sistemes de mesura de reputació, tal i com comenta l'autor a les conclusions de l'article.

També a les conclusions es planteja instal·lar varis nodes de captura de tuits, per tal d'assegurar-

169 Bailey, Fenn. "fennb/phirehose." GitHub. <https://github.com/fennb/phirehose/wiki/Introduction> (consultat a 11 de maig de 2014)

170 Bailey, Fenn. "fennb/phirehose." GitHub. <https://github.com/fennb/phirehose/wiki/Introduction> (consultat a 11 de maig de 2014) pg. 14

se que no se'n perd cap durant el procés. És una bona estratègia però fa encara més complexe que qualsevol persona que no tingui coneixements aplicats d'aquests sistemes informàtics pugui dur-ho a terme.

El treball de Catalina Oliver Vidal "Ontologia de Twitter : El primer pas per inferir informació dels tweets"¹⁷¹, també ofereix un bon anàlisi de les possibilitats de l'API de Twitter tal i com ho podria encaixar en la meua concepció, en concret per crear tota una ontologia. I s'aprofita del programa Protégé¹⁷² i el format owl¹⁷³. Així va poder dissenyar una ontologia bàsica amb les classes i propietats que es necessiten per poder guardar el contingut dels tuits de Twitter i relacionar-los amb els Usuaris i amb els continguts del propi tuit.

Catalina Oliver ens explica amb gran encert també que si volem entendre la web semàntica hem de disposar d'aquestes ontologies, les quals permeten explicar de manera molt visual i gràfica les entitats i com es coordinen els conceptes, les relacions, les instàncies i els atributs. Generar mapes d'entitats, les quals estructuraran el llenguatge més humà que farem entendre a les màquines, una estructura conceptual que sigui compartida, processada i interpretada per agents de programari.

En el seu treball exposa molt clarament en taules les definicions d'algunes de les entitats, com nosaltres també havíem fet a l'apartat "Explicar més coses amb les metadades JSON". Dades que podem obtenir senzillament fent una ullada a la web de Documentació de Twitter¹⁷⁴. Ella va desglossant des de l'apartat d'usuaris, a la de tuits, a la de hashtags, continguts multimèdia,... I un cop ha reconegut totes les entitats, les va processar amb el Protégé per elaborar els diagrames amb les relacions entre entitats.

El que potser sorprèn és què en fa després amb la ontologia i és voler-ne elaborar una espècie d'entorn web amb el programari Netbeans, les llibreries JENA, el servidor web Apache Tomcat, una màquina virtual amb Virtualbox i Windows XP¹⁷⁵. Però no sabem quin ús se'n podria fer ja que el treball acaba sobtat amb unes conclusions molt breus i cordials.

3.4.1 El model OAIS i iArxiu

No hem volgut acabar l'apartat més personal de la investigació del treball sense provar de

171 Catalina, Oliver Vidal. "Ontologia de Twitter : El primer pas per inferir informació dels tweets." Repositori institucional - UOC. <http://openaccess.uoc.edu/webapps/o2/handle/10609/28324> [<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/28324/6/coliverTFC0114memoria.pdf>] (consultat a 11 de maig de 2014)

172 "A free, open-source ontology editor and framework for building intelligent systems." protégé. <http://protege.stanford.edu> (consultat a 11 de maig de 2014)

173 Col·laboradors de Viquipèdia, "OWL," Viquipèdia, l'Enciclopèdia Lliure, //ca.wikipedia.org/w/index.php?title=OWL&oldid=13278027 (descarregat 26 d'abril de 2014)

174 "Documentation." Twitter Developers. <https://dev.twitter.com/docs> (consultat a 11 de maig de 2014)

175 "El 8 de abril de 2014 finaliza el Soporte. Fin de soporte Windows XP." <http://www.microsoft.com/es-es/windows/endofsupport.aspx> (consultat a 14 de maig de 2014) Suposem que amb la fi del suport a aquest sistema operatiu l'autora de l'article ja en deu descartar l'ús.

portar a la pràctica el treball amb documentació de Twitter a través d'eines pròpies de l'Arxivística. És per això que hem hagut de provar com encaixaria el model OAIS i el sistema iArxiu en tot plegat.

Si considerem el model OAIS d'arxiu digital¹⁷⁶ en que hem de tenir un paquet d'informació format per la informació del contingut i la informació sobre el contingut, podríem pensar que el que ens apareix al navegador és una imatge gràfica de la informació del contingut i llavors totes les metadades que rebem quan fem peticions a l'API formarien la informació sobre la preservació.

En les tres variants informatives¹⁷⁷, l'enviament o adquisició seria la informació que s'aporta al enviar (que podem consultar a la pàgina corresponent a la API), la informació d'arxiu o de conservació seria com acaba integrada la informació a la base de dades MySQL, i la informació de difusió o de recuperació seria les respostes de l'API de Twitter o bé l'arxiu en format CSV que rebem en les opcions de Twitter.

La pròpia base de dades MySQL de Twitter seria el document complet, la unitat documental composta que anomenàvem a l'inici del treball. Aquesta base de dades amb taules de tuits de cada usuari, és el que segurament rebrà la Library of Congress. Nosaltres només en rebem per cada usuari un extracte que és aquest CSV on rebem dades estructurades, tal i com correspondria a una base de dades.¹⁷⁸ A més, obtenim una base de dades tipus font ja que la pròpia base de dades conté els documents, no pas referències a documentació externa.¹⁷⁹

Doncs bé, podem considerar com es valoraria l'ingrés d'aquest tipus de documentació en un sistema iArxiu. Així doncs el primer que farem serà elaborar un esquema de metadades a partir del fitxer CSV que ens ofereix Twitter, feina que s'assemblaria molt a la que exposava Catalina Oliver amb les ontologies. Per fer-ho simple Twitter només afegiríem els següents camps (o identitats) que es corresponen als que venen amb el fitxer CSV:

- tweet_id
- in_reply_to_status_id
- in_reply_to_user_id
- timestamp
- source

176 Col·laboradors de Viquipèdia, "Model de preservació OAIS," Viquipèdia, l'Enciclopèdia Lliure, [//ca.wikipedia.org/w/index.php?title=Model_de_preservaci%C3%B3_OAIS&oldid=13346769](http://ca.wikipedia.org/w/index.php?title=Model_de_preservaci%C3%B3_OAIS&oldid=13346769) (descarregat 7 de maig de 2014)


177 "Modelo OAIS (Open Archival Information Systems)." Docusystem. <http://docusystem.blogspot.com.es/2006/11/modelo-oais.html> (consultat a 11 de maig de 2014)

178 SOLER, Joan. *Del bit al logos. Preservar documents electrònics a l'Administració local*. Col·lecció Estudis. Diputació de Barcelona, octubre de 2009, Barcelona. pg. 26

179 SOLER, Joan. *Del bit al logos. Preservar documents electrònics a l'Administració local*. Col·lecció Estudis. Diputació de Barcelona, octubre de 2009, Barcelona. pg. 29

- text
- retweeted_status_id
- retweeted_status_user_id
- retweeted_status_timestamp
- expanded_urls

Podríem crear un Vocabulari anomenat “CSV Twitter”.



Usuari Formació 01
 Arxiver , Administrador de fons
 ESAGED_2013/Aniol Maria Vallès
[Canvia ens/fons](#) | [Surt](#)

Inici
Ingrés
Consulta
Administració

Ets a: > [Inici](#) > [Administració](#) > [Gestió vocabularis](#) > [Creació vocabularis](#)

Menú administració

- Gestió sèries documentals
- Gestió polítiques d'accés
- Gestió polítiques de disposició
- Gestió polítiques de preservació
- Gestió d'usuaris
 - Vista usuaris-grups
 - Vista grups-usuaris
 - Assignació usuaris arxivats
- Gestió plantilles
 - Creació plantilles
 - Administració plantilles
- Gestió vocabularis
 - Creació vocabularis
 - Consulta vocabularis
- Gestió carpetes
- Consulta registre d'accions
- Consulta registre d'accessos
- Consulta estadístiques

Creació vocabularis

Configuració vocabulari

Nom vocabulari:

[Afegeix element](#) |
 [Modifica element](#) |
 [Elimina element](#)

	Nom	Tipus	Valor per defecte	Llista de valors	Valor fixe	Obligatori
<input checked="" type="radio"/>	tweet_id	Text				Si
<input type="radio"/>	in_reply_to_status_id	Text				No
<input type="radio"/>	in_reply_to_user_id	Text				No
<input type="radio"/>	timestamp	Data/hora				Si
<input type="radio"/>	source	Text				Si
<input type="radio"/>	text	Text				Si
<input type="radio"/>	retweeted_status_id	Nombre real				No
<input type="radio"/>	retweeted_status_timestamp	Data/hora				No
<input type="radio"/>	expanded_urls	Text				No

Tot seguit crearem una plantilla també a l'iArxiu.

Per a provar el sistema d'iArxiu vaig optar per crear un usuari de Twitter que automàticament agafava enllaços de varis fils RSS i en publicava les noves publicacions en tuits a Twitter. El vaig anomenar @usproiarx¹⁸⁰, i com a fils RSS li vaig assignar els de:

- Featured pictures on Wikimedia Commons – commons.wikimedia.org
http://tools.wmflabs.org/catfood/catfood.php?category=Featured_pictures_on_Wikimedia_Commons
- Archive.org Most Recent Audio
<https://archive.org/services/collection-rss.php?mediatype=audio>
- Archive.org Most Recent Moving Image
<https://archive.org/services/collection-rss.php?mediatype=movies>
- The Creative Commons (pls read the rules) Pool, with geodata
<https://api.flickr.com/services/feeds/geo/?g=26241990@N00>

Com a sistema per revisar i publicar tuits vaig optar per Dlvr.it¹⁸¹ ja que permet d'una manera senzilla i força adaptada a Twitter, automatitzar la publicació de tuits sense opcions massa complexes. I el resultat van ser més de 400 tuits en un dia.



180 No sé fins quan el mantindré actiu a Twitter, o fins quan Twitter el tolerarà com a bot automàtic que tan sols publica tuits amb referències a continguts que obté de fonts RSS i sense cap interacció de menció dins de Twitter amb d'altres usuaris.

181 "dlvr.it." dlvr.it blog RSS. <https://dlvr.it> (consultat a 15 de juny de 2014)

Doncs bé, seguint amb la prova, vaig donar d'alta una sèrie per als arxius CSV de Twitter.

Alta sèrie documental

* Camps obligatoris.

* Codi de classificació:	<input type="text" value="TW_0001"/>
* Títol:	<input type="text" value="CSV_Twitter"/>
Descripció:	<input type="text" value="csv amb tuits de Twitter"/>
Nivell de descripció :	<input type="text" value="Sèrie"/>
Nom productor:	<input type="text" value="Twitter"/>
Codi intern:	<input type="text"/>
Data inici:	<input type="text" value="21/05/2014"/>  ▼
Data final:	<input type="text"/>  ▼

I vaig passar a ingressar el CSV a iArxiu omplint-ne les meta-dades d'exemple. I llavors en vaig rebre el Paquet d'Informació de Transferència.







```
<?xml version="1.0" encoding="UTF-8"?>  
<mets:mets TYPE="urn:iaxii:2.0:templates:catcert:PL_document_demo" xmlns:mets="http://www.loc.gov/METS/"  
><mets:dmdSec ID="DMD_1"><mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="urn:iaxii:2.0:vocabularies:catcert:Voc_document"  
MIMETYPE="text/xml"><mets:xmlData><voc:document xmlns:voc="http://schemas.user.iaxii.org/com/2.0/Voc_document"  
><voc:codi_referencia>CAT/ESAGED</voc:codi_referencia><voc:codi_classificacio>TW_0001  
</voc:codi_classificacio><voc:títol_serie_documental>CSV_Twitter</voc:títol_serie_documental><voc:títol>Sol·licitud  
alta xxxx</voc:títol><voc:data_creacio>2014-05-21T18:08:00</voc:data_creacio><voc:nivell_descripcio>Unitat  
documental composta</voc:nivell_descripcio><voc:suport>electrònic</voc:suport><voc:nom_productor>ESAGED  
</voc:nom_productor><voc:típus_document>Registre</voc:típus_document><voc:classificacio_seguretat>Accés públic  
</voc:classificacio_seguretat_acces><voc:sensibilitat_dades_LOPD>Nivell baix  
</voc:sensibilitat_dades_LOPD><voc:nivell_classificacio_evidencial>Evidència d'origen de dades  
</voc:nivell_classificacio_evidencial><voc:document_essencial>No  
</voc:document_essencial></voc:document></mets:xmlData></mets:mdWrap></mets:dmdSec><mets:fileSec><mets:fileGrp ID=  
"BIN_1"><mets:file ID="BIN_1.0" MIMETYPE="application/msaccess" CHECKSUM="fc4fa9598fdb10d4e279068e989f1871af6e7563"  
CHECKSUMTYPE="SHA-1" CREATED="2014-05-21T18:04:35.483+02:00"><mets:FContent><mets:binData>  
InR3ZWVOX2LkIiwiaW5fmGd9f3cRhDHvZ21kIiwiaw5fcmVwbHlfG9fdXNlcl9pZCIsInRpbWVzdGFtcCIisInnVvdXJjZSIsInRleHQiLCJyZy  
XR3ZWVOZWRf3cRhDHvZ21kIiwicmVod2VldGVkX3NOYXRlc19lc2VyZ21kIiwicmVod2VldGVkX3NOYXRlc19oaWllc3RhbXAiLCJleHBhbmRlZF91c
```

Aquest podria ser un exemple de com amb l'esquema de metadades METS¹⁸² es podria resoldre un paquet del fitxer CSV.






182 Col·laboradors de Viquipèdia, "Estàndard de Codificació i Transmissió de Metadades," Viquipèdia, l'Enciclopèdia Lliure, //ca.wikipedia.org/w/index.php?title=Est%C3%A0ndard_de_Codificaci%C3%B3_i_Transmissi%C3%B3_de_Metadades&oldid=12809392 (descarregat 19 de gener de 2014)

Després vaig fer una prova d'ingrés del fitxer CSV en format d'expedient (en ZIP).

document

* Codi de referència:	<input type="text" value="TW_UsProiArx_0001"/>
Número de document:	<input type="text"/>
* Títol:	<input type="text" value="Tweets d'UsProiArx"/>
* Data de creació:	<input type="text" value="22/05/2014 11:26"/> 
* Nivell de descripció:	<input type="text" value="Unitat documental composta"/>
* Volum i Suport:	<input type="text" value="423 tuits"/>
Nom del productor:	<input type="text" value="Twitter"/>
Unitat productora:	<input type="text" value="UsProiArx"/>
Descripció:	<input type="text" value="sProiArx, l'usuari de prova amb feeds RSS."/>
Descriptors:	<input type="text"/>
Tipus de document:	<input type="text" value="Registre"/> 
Classificació seguretat i accés:	<input type="text" value="Accés públic"/> 
Sensibilitat dades LOPD:	<input type="text" value="Nivell baix"/> 
Nivell de classificació evidencial:	<input type="text" value="Evidència d'origen de dades"/> 
Document essencial:	<input type="text" value="No"/> 

expedient

* Codi de referència:	<input type="text" value="TW_UsProiArx_0001"/>
* Número expedient:	<input type="text" value="0001"/>
* Codi de classificació:	<input type="text" value="TW_0001"/>
* Títol sèrie documental:	<input type="text" value="CSV_Twitter"/> 
* Nivell de descripció:	<input type="text" value="Unitat documental composta"/>
* Títol:	<input type="text" value="Tweets d'UsProiArx"/>
* Data d'obertura:	<input type="text" value="21/05/2014 11:31"/> 
* Data de tancament:	<input type="text" value="22/05/2014 11:31"/> 
* Nom del productor:	<input type="text" value="Twitter"/>
Unitat productora:	<input type="text" value="UsProiArx"/>
Descripció:	<input type="text" value="sProiArx, l'usuari de prova amb feeds RSS."/>
Descriptors:	<input type="text"/>
Documentació relacionada:	<input type="text"/>
Tipus de relació:	<input type="text"/>
* Classificació seguretat i accés:	<input type="text" value="Accés públic"/> 
Sensibilitat dades LOPD:	<input type="text" value="Nivell baix"/> 

Però no va acabar de quallar ja que no vaig aconseguir que el CSV acabés reconegut com un expedient que conté tuits. El Vocabulari de metadades tot i que permetia configurar bastant els camps que sí apareixien al fitxer CSV, no vaig assolir que es relacionin.

En una segona prova vaig poder introduir una nova plantilla de document i descriure ja sí el fitxer CSV a nivell d'expedient.

expedient

* Codi de referència:	<input type="text" value="TW_UsProiArx_0001"/>
* Número expedient:	<input type="text" value="0001"/>
* Codi de classificació:	<input type="text" value="TW_0001"/>
* Títol sèrie documental:	<input type="text" value="CSV_Twitter"/>
* Nivell de descripció:	<input type="text" value="Unitat documental composta"/>
* Títol:	<input type="text" value="Tweets d'UsProiArx"/>
* Data d'obertura:	<input type="text" value="21/05/2014 10:16"/>
* Data de tancament:	<input type="text" value="22/05/2014 10:16"/>
* Nom del productor:	<input type="text" value="Twitter"/>
Unitat productora:	<input type="text" value="UsProiArx"/>
Descripció:	<input type="text" value="Tweets d'UsProiArx, l'usuari de prova amb"/>
Descriptors:	<input type="text"/>
Documentació relacionada:	<input type="text"/>
Tipus de relació:	<input type="text"/>
* Classificació seguretat i accés:	<input type="text" value="Accés públic"/>
Sensibilitat dades LOPD:	<input type="text" value="Nivell baix"/>

El problema però en aquest nou intent és que iArxiu no ens accedia al contingut del CSV i per tant no podíem descriure cada registre o tuit que es contenia en el fitxer.

Fins aquí va ser la nostra prova amb la plataforma iArxiu. No podia anar més enllà amb iArxiu ja que estava tractant amb un sistema datacèntric (com diria Jordi Serra), per tant no podia plantejar la separació en documents simples que tampoc sabia en quin format es podrien mantenir estables ja que perdríem el conjunt lògic de la documentació que tractem. Així que com a prova vaig decidir deixar-ho aquí i no marejar més el tema.

Podria també elucubrar sobre com seria un model de metadades especificat per la documentació de Twitter. Per exemple començant per l'Encoded Archival Description (EAD) que és un DTD creat amb SGML entre els anys 1995 i 1999 en l'àmbit anglosaxó i coordinat per l'Encoded Archival Description Working Group de la Societat d'Arxivers nord-americans,¹⁸³ i que presumiblement serà el model que adopti la Library of Congress. L'esquema de descripció llavors

¹⁸³ SOLER, "Joan. Del bit al logos. Preservar documents electrònics a l'Administració local". Col·lecció Estudis. Diputació de Barcelona, octubre de 2009, Barcelona. pg. 125

seria el de:

- Element name
- Description
- May contain
- May occur within
- Attributes
- Example

I dins de cada apartat els atributs (*Attributes*):

- ID (si ha de portar una identificació)
- IDREF (si fa referència a una altra ID)
- LANGCODE (si ha de portar un codi d'idioma)

També trobaríem els valors per cada atribut:

- #IMPLIED l'atribut no està requerit.
- #REQUIRED sí que està requerit.
- #FIXED l'atribut tan sols té una opció de valor possible.

No sé però quins Elements triarà la Library of Congress però ens en podríem fer una idea consultant-ne el llistat¹⁸⁴ i comparant-lo amb algunes de les metadades que ens exposa Twitter¹⁸⁵, i també amb la Dublin Core¹⁸⁶. També he provat de trobar-ne alguna relació amb PREMIS¹⁸⁷ però no amb massa encert:

- **<accessrestrict> Conditions Governing Access:** Indica si hi ha algun element de copyright. També podria ser el **<legalstatus> Legal Status** o el **<userrestrict> Conditions Governing Use**. I podria relacionar-se amb *withheld_copyright* de Twitter que està orientat també a indicar aquesta mena de restriccions i estatus legals. A Dublin Core es correspondria amb *rights*. I a PREMIS seria potser *rightsStatement*.

184 Society of American Archivists. "EAD Tag Library." EAD: Encoded Archival Description Version 2002 Official Site (EAD Official Site, Library of Congress).

http://www2.archivists.org/sites/all/files/EAD2002TL_5-03-V2.pdf (consultat a 8 de juny de 2014) pg. 23

185 "Tweets." Twitter Developers. <https://dev.twitter.com/docs/platform-objects/tweets> (consultat a 8 de juny de 2014)

186 "Dublin Core Metadata Element Set, Version 1.1." DCMI Home: Dublin Core© Metadata Initiative (DCMI). <http://dublincore.org/documents/dces/> (consultat a 8 de juny de 2014)

187 Library of Congress. "PREMIS Data Dictionary for Preservation Metadata, Version 2.2 (July 2102) - premis-2-2.pdf." PREMIS: Preservation Metadata Maintenance Activity (Library of Congress). <http://www.loc.gov/standards/premis/v2/premis-2-2.pdf> (consultat a 9 de juny de 2014)

- **<archdesc> Archival Description:** Per descriure contingut extra que no aparegui en el tuit al fer-ne una descripció més extensa. A PREMIS seria *eventDetail*.
- **<author> Author:** Indicaria l'autor del tuit. Es relacionaria amb *user* de Twitter, i per tant hauria d'indicar moltes més metadades que sí exposa Twitter, com la identificació numèrica de l'usuari, el seu sobrenom, etcètera. En la DublinCore correspondria a *creator*. I PREMIS seria *linkingAgentIdentifier* o *linkingAgentIdentifierValue* per a la sèrie alfanumèrica de Twitter. Tant a Twitter com a Dublin Core hi ha la metadada *contributor* (a PREMIS seria *linkingAgentRole*), però tal i com hem explicat a Twitter encara no se li ha donat cap utilitat pública.
- **<date> Date:** Per la data del tuit. A Twitter correspondria amb *created_at*. A Dublin Core seria també *date*. I a PREMIS seria *eventDateTime*.
- **<language> Language:** Idioma del contingut en codi ISO 639-2b (o ISO 15924 si s'especifica) de 3 lletres. A Twitter es correspondria amb *lang* i s'hauria de mirar la correspondència de codificació ja que a Twitter és BCP 47 que està basat en la ISO 639. En el cas de Dublin Core es correspondria amb *language*.
- **<note> Note:** Podria oferir un espai per a anotacions extra pels tuits.
- **<resource> Resource:** Podria fer referència als elements extra que s'adjunten al tuit, el que a Twitter s'anomenen les *entities*¹⁸⁸.
- **<unitid> ID:** La sèrie alfanumèrica d'identificació de la unitat que es descriu. A Twitter es correspondria també amb *id*. A Dublin Core es correspondria amb *identifier*. I a PREMIS seria *objectIdentifier*.

Amb el llenguatge de marcatge METS no hem trobat cap perfil publicat que ens pogués servir massa d'exemple¹⁸⁹, així que ens remetem com a exemple el que ens oferia ja la plataforma iArxiu després de l'ingrés del paquet d'informació del fitxer CSV.

Però tot plegat no són més que elucubracions. El fet és que sense més indicacions per part de la Library of Congress no podem saber quina serà la forma definitiva de com es preservaran els tuits de Twitter. No sabem si es farà també una preservació de les fulles d'estil CSS¹⁹⁰ per tal d'oferir quelcom més que el contingut textual dels tuits. Tampoc sabem quina preservació rebrà la base de dades MySQL si és que serà en format d'exportació tal i com s'està transferint tot el

188 "Entities." Twitter Developers. <https://dev.twitter.com/docs/platform-objects/entities> (consultat a 8 de juny de 2014)

189 "METS Example Documents." Metadata Encoding and Transmission Standard (METS) Official Web Site. <http://www.loc.gov/standards/mets/mets-examples.html> (consultat a 8 de juny de 2014)

190 Col·laboradors de Viquipèdia, "Cascading Style Sheets," Viquipèdia, l'Enciclopèdia Lliure, http://ca.wikipedia.org/w/index.php?title=Cascading_Style_Sheets&oldid=13549168 (descarregat 9 de juny de 2014)

conjunt de tuits.

O si tot quedarà només en una exposició de les metadades en forma d'exportació en CSV, JSON o XML com hem vist en el cas dels National Archives del Regne Unit.

És tot molt incipient i tan sols n'hem pogut fer tot aquest plegat d'especulacions per la nostra banda ja que ni la Library of Congress ni Twitter tenen pensat fer cap més comunicat informatiu fins que no s'ofereixi la col·lecció al públic.

4 Conclusions

Podria parlar de les coses que he après fent aquest treball final. La primera d'elles és la de conèixer molt més sobre el funcionament de les API i la rellevància que tindran amb l'onada aquesta que ens arriba de fal·lera per la obertura de dades, ja que l'accés a aquestes dades es farà per API. He après a treballar amb Postman, cosa que m'ha servit com a punt de partida per poder pensar en com explotar d'altres APIs d'altre serveis web.

Una altra cosa que he après és que malgrat s'accedeixi a les dades en JSON, XML o el que se m'ofereixi, amb Twitter només accediré a les dades que se'm vulguin oferir i no pas obligatòriament a totes les que existeixin en la base de dades ni tampoc tindrè els documents originals íntegres. Per a tal tasca hauria d'accedir a les bases de dades per exemple on m'apareixerien totes les dades en taules.

4.1 Objectius i preguntes

Ens vam proposar varis objectius per aquest treball:

- **Explicar perquè entenem que Twitter pot ser tractat arxivísticament.** Sí, fins a on hem pogut hem intentat explicar la profunditat que podria tenir tractar Twitter arxivísticament (treball amb l'API i les metadades) i quines limitacions podríem trobar-nos (termes legals i tècnics del servei).
- **Exposar quines eines ofereix Twitter per al seu anàlisi i treball per part dels arxivers.** També ho hem pogut assolir bastant, mostrant les diverses maneres com s'havia tractat la feina amb Twitter, des dels que es dedicaven sols a fotografiar amb captures de pantalla fins els que bolcaven codi en suports d'emmagatzematge.
- **Arribar a establir uns paràmetres i concepcions que ajudin a gestionar aquesta eina de comunicació dels arxius que pot ser Twitter.** A l'apartat "2.4. L'estudi de Twitter" (i sobretot al subapartat "Implicació a Twitter"), hem exposat diverses concepcions de com s'ha plantejat la gestió d'eines de comunicació dels arxius a Twitter.
- **Definició de l'objecte digital, explotació de l'objecte digital.** Hem explicat bastant com

hauríem d'entendre Twitter, la concepció que hem de tenir de les dades emmagatzemades en una base de dades MySQL, com aquesta se'ns mostra en PHP, i com la rebem en un fitxer CSV o bé en codi JSON i XML.

- **Plantejar un model de preservació de tuits que no sigui simples captures de pantalla com fa Padicat.** A "2. Casos arxivístics amb Twitter" i sobretot al subapartat "Quina podria ser la meua proposta", hem provat d'exposar aquest model malgrat que a iArxiu no ens n'hàgim sortit massa.

I ens van sorgir vàries preguntes:

- **Estarem preservant els tuits de Twitter en una col·lecció però aïllant-los del seu context a Internet?** Podria ser. Com estàvem concloent a l'anterior apartat teòric, desconeixem quin tractament rebrà la resta d'elements de Twitter, ja que segons el contracte de cessió només són els tuits i la informació que aporten el que es preservarà a la Library of Congress.
- **Quin interès tan gran hi ha en descarregar tuits de Twitter si la gràcia és assegurar-se que aquests estiguin a Internet i no pas en un suport fora de la xarxa?** Se m'ha fet molt curiós veure com malgrat que descarregar i arxivar tuits massivament va contra els Termes de servei de Twitter actuals¹⁹¹, apareixen molts serveis que permeten fer-ho. Segurament s'aprofiten de la opció que dona Twitter d'obtenir un CSV amb els nostres tuits i de sobreexplotar l'API de Twitter. Ens sembla si més no curiós, i caldrà veure com evoluciona sobretot amb el fenomen aquest de les infografies amb Big Data i com s'hi posiciona finalment Twitter.

4.2 Recorregut del treball

Al llarg d'aquest treball he analitzat els casos en que s'ha volgut treballar amb Twitter i puntualitzar sobre quins han anat directament a les entranyes dels documents anomenats tuits i quins s'han quedat en l'aparença externa. També quins han elaborat estudis més complexes i quins s'han quedat en les estadístiques. Ha sigut interessant per veure com amb el temps es van adoptant els coneixements més útils i adequats per tractar amb tota la seva complexitat l'entorn documental de Twitter. I com es va perdent la por a noves tecnologies i llenguatges que porten l'Arxivística a un nivell més profund de la informàtica i la documentació electrònica.

He pogut saber que la Library of Congress va rebre en el seu moment de Twitter uns 21 bilions de tuits amb prop de 50 camps de meta-dades (com hem indicat a l'apartat "L'exemple de NARA"). Per tant podríem imaginar-nos la magnitud del que pot representar el projecte final comparat amb el que nosaltres hem pogut treballar d'un CSV que tan sols contenia 10 camps.

¹⁹¹ Terms of Services (ToS) (Section 4A) "Developer Rules of the Road." Twitter Developers.
<https://dev.twitter.com/terms/api-terms> (consultat a 12 de juny de 2014)

I finalment haurem d'estar atents també al que es publiqui a través del programa "Twitter Data Grants" que he anomenat a l'apartat de "Quina podria ser la meva proposta".

Podria fer un repas encara més detallat de tot el que he descrit i avaluat al llarg del treball però cada tema ja té el seu apartat propi i la valoració relacionada i no voldria fer-me pesat.

4.3 Projectes a dur a terme

Com a projectes a partir d'aquest treball me'n podria plantejar varis. Per una banda aprendre molt més sobre l'ús d'eines com les API, cosa que em comportaria plantejar-me d'una vegada aprendre llenguatges de programació destacats com Python o almenys PHP i Javascript. Per altra banda un projecte posterior seria utilitzar els coneixements que adquirís de programació per programar i publicar una aplicació que permetés facilitar encara més la consulta de les metadades de Twitter, seguint bastant l'estil d'aplicacions web com FOCA¹⁹². Aquesta aplicació agafaria cada tipus de consulta que se li pot fer a l'API de Twitter i a través d'un senzill formulari oferir als usuaris l'obtenció de les metadades d'un tuit, d'un perfil o d'una cerca, sense haver de descarregar-se cap navegador com el Google Chrome ni instal·lar-se cap extensió com Postman. La problemàtica vindria amb com evitar emmagatzemar les dades sensibles que requereix l'autenticació identificativa dels usuaris que volen fer ús de l'API, però és un aspecte que ja s'haurà de resoldre si algun dia es planteja en ferm el projecte.

4.4 Agraïments

Al Joan Soler @Diplomaticat per acceptar des del primer dia que pogués dur a terme aquest Treball Final de Màster, fins i tot abans que fos formalment plantejat.

A la família que m'ha aguantat aquests dos anys de viatges amunt i avall, sense entendre massa què coi m'importava tant d'estudiar allà l'ESAGED.

Al Jordi Lopez @wuyingren per ajudar-me a entendre com podia interactuar de manera molt amigable (dintre el que permet el sistema) amb la API de Twitter.

Al Raimon Nualart @raimonnualart per atendre'm les consultes sobre l'iArxiu.

Als amics de Reus i d'arreu, que cada cap de setmana i qualsevol dia de l'any ens acompanyaríem sense importar quina fos l'hora que ens n'anem a dormir (però anem a dormir).

I finalment a qui segueixo i qui em segueix a Twitter i que em permeten descobrir coses que jo per mi sol no havia conegut encara.

192 "FOCA Online - Informática 64." FOCA Online - Informática 64. <http://www.informatica64.com/foca/default.aspx> (consultat a 12 de juny de 2014)

5 Glossari

API (*Application Programming Interface*): Interfície d'interacció amb un servei web. Conjunt d'ordres i funcions que permeten comunicar-se amb un sistema i que ens respongui. S'acostuma a oferir també el conjunt de llibreries, protocols, i llenguatges amb els quals es pot accedir al seu funcionament. Les declaracions que definiran el contracte d'un component informàtic amb qui farà ús dels seus serveis.

Aplicacions natives: Aplicacions o programes informàtics que s'executen en el propi aparell de l'usuari. Es contraposa a les aplicacions al núvol o entorn virtual.

Base 64: Sistema de numeració que permet el seu ús per la codificació de documents electrònics i poder-los incrustar dins un codi.

BCP 47 (identificació d'idioma): És un codi de llenguatge estàndard de l'IETF (*Internet Engineering Task Force*) que permet identificar de forma estructurada i jerarquitzada les llengües, famílies i varietats lingüístiques. En català per exemple diferenciaria *ca_ES* pel català parlat en l'Estat Espanyol del *ca_AD* que seria el parlat a Andorra.

Còpies locals: Les duplicacions que es fan d'uns fitxers al suport d'emmagatzematge del propi aparell electrònic que dirigeix l'usuari.

CSV: És un tipus de document en format obert senzill per representar dades en forma de taula, en que les columnes se separen per comes i les files per salts de línia. Els camps que continguin una coma, un salt de línia o una cometa doble han de ser tancats entre cometes dobles.¹⁹³

Data centers: Grans centres de processament de dades, que treballen amb sistemes ingents de servidors i sistemes de còpies redundants, per tal d'emmagatzemar el màxim possible de dades i mantenir-les el més incorruptes possible.

Direct Message (DM): Missatge privat entre usuaris de Twitter. Queda fora de tota visibilitat per part de la resta d'usuaris, ja que només es permet enviar aquests missatges a un sol usuari i de cap manera a varis.

DMCA (*Digital Millennium Copyright Act*): És una llei de copyright nord-americana que començà a ser efectiva el 28 d'octubre de 1998 i que criminalitza la producció i disseminació de tecnologies, dispositius o serveis que permetin burlar el control d'accés a obres amb copyright. Amb aquesta llei es pot perseguir a qui distribueixi material sota copyright sense els permisos legals, i demanar-ne la supressió de la difusió.¹⁹⁴

¹⁹³ Col·laboradors de Viquipèdia, "CSV," Viquipèdia, l'Enciclopèdia Lliure, [//ca.wikipedia.org/w/index.php?title=CSV&oldid=13096365](http://ca.wikipedia.org/w/index.php?title=CSV&oldid=13096365) (descarregat 20 de març de 2014)

¹⁹⁴ Wikipedia contributors, "Digital Millennium Copyright Act," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Digital_Millennium_Copyright_Act&oldid=608434500 (consultat a 16 de maig de 2014)

Entitats (API Informàtica): Els varis camps amb informació que apareixen en una consulta a la API. Poden venir definides i acotades segons les variables que s'hagin afegit a les peticions i ordres a la API.

Followers: Usuaris seguidors d'un perfil a Twitter.

Followings: Perfils que està seguint un usuari concret de Twitter.

Gift Agreement (*The Deed of Gift*): Formulari legal que permet l'acceptació d'una transferència entre un donant i un dipòsit. Transfereix la propietat i els drets legals dels materials donats. El document un cop signat estableix les bases legals entre el donant i el dipòsit, i l'estatus legal dels materials dipositats.¹⁹⁵ Té molta similitud amb els "contractes de cessió de dades".

Hash: És el resum de l'algorisme o funció per resumir i identificar una dada a través de la probabilitat. Es pot obtenir amb l'algorisme MD5, SHA-1 o CRC.

Hashtag: Etiqueta temàtica que es pot introduir en els tuits i els missatges directes de Twitter. Ve precedida per un coixinet # davant de la paraula per indicar-la.

Internet de les coses: Evolució de la xarxa d'Internet orientada a connectar tots els aparells que ens envolten i a que es puguin comunicar i fer-los interactuar entre ells¹⁹⁶. Ha guanyat difusió i popularitat des de l'aparició dels *gadgets* i aparells intel·ligents com els *smartphones* o els *weareables*. Les ontologies s'espera que ajudin com a eina per a la seva implantació.

JSON (*JavaScript Object Notation*): És un llenguatge de marcatge estàndard obert basat en text dissenyat per a intercanviar dades llegibles per humans. Deriva del llenguatge Javascript, i serveix per representar estructures de dades simples i llistes associatives o d'objectes.¹⁹⁷

Microblog: Format de blog en que les publicacions han de ser el més breus possibles. Es va triar l'estàndard de 140 caràcters per mimetisme amb els SMS. El servei web més conegut dels quals és Twitter, però n'han existit molts d'altres, i fins i tot va existir el Google Buzz que era l'alternativa que volia oferir Google però que no va durar gaire més d'un any¹⁹⁸.

MySQL: Sistema de bases de dades de programari lliure i codi obert, que treballa amb SQL. Actualment és propietat d'Oracle.

OAIS (*Open Archival Information System*): És un model de preservació. Té una doble finalitat:

¹⁹⁵ "A Guide to Deeds of Gift." Society of American Archivists.

<http://www2.archivists.org/publications/brochures/deeds-of-gift> (consultat a 16 de maig de 2014)

¹⁹⁶ Col·laboradors de Viquipèdia, "Internet de les coses," Viquipèdia, l'Enciclopèdia Lliure, [//ca.wikipedia.org/w/index.php?title=Internet_de_les_coses&oldid=13308812](http://ca.wikipedia.org/w/index.php?title=Internet_de_les_coses&oldid=13308812) (descarregat 30 d'abril de 2014)

¹⁹⁷ Col·laboradors de Viquipèdia, "JSON," Viquipèdia, l'Enciclopèdia Lliure, [//ca.wikipedia.org/w/index.php?title=JSON&oldid=12835673](http://ca.wikipedia.org/w/index.php?title=JSON&oldid=12835673) (descarregat 20 de gener de 2014)

¹⁹⁸ Wikipedia contributors, "Google Buzz," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Google_Buzz&oldid=602256102 (consultat a 16 de maig de 2014)

la preservació i l'accés a llarg termini de la informació en format digital.¹⁹⁹

OAuth: És un protocol obert proposat per Blaine Cook i Chris Messina que permet l'autorització segura d'una API d'una manera estàndard. Aquest sistema d'identificació va ser utilitzat per Twitter per facilitar la interacció dels seus usuaris amb d'altres serveis web que treballen amb l'entorn i les dades de Twitter. Es basa en el model que va establir OpenID en el qual el registre d'usuaris es fa des d'una sola web, i després es crea una passarel·la segura per a facilitar el registre a d'altres serveis web sense haver d'aportar excessives dades identificatives com una contrasenya.

Ontologia (Informàtica): Elaboració exhaustiva i rigorosa d'un esquema conceptual dins d'un o varis dominis, amb la finalitat de facilitar la comunicació i l'intercanvi d'informació entre diferents sistemes i entitats.

OpenID: Estàndard d'identificació digital descentralitzat, amb el que l'usuari pot identificar-se en una pàgina web a través d'una URL i permet ser verificat per qualsevol servidor que suporti aquest protocol. Així els usuaris no es veuen obligats a crear-se un nou compte d'usuari per obtenir accés i tan sols se'ls requereix d'un identificador creat en un servidor que verifiqui OpenID.

Owl (Format informàtic): El *Web Ontology Language* és un llenguatge dissenyat per ser usat en aplicacions que necessitin processar el contingut i informació enlloc de presentar-lo només com a informació per a humans²⁰⁰, autoritzant ontologies o coneixements per a les bases de dades. Ofereix més facilitat d'interpretació del contingut web que XML o RDF. Té el suport del consorci W3C (*World Wide Web Consortium*), i ha atret l'atenció tant del món acadèmic, com del mèdic i comercial.²⁰¹

PHP: Llenguatge de programació interpretat al servidor web i que genera pàgines web dinàmiques. Al executar-se al servidor aquest últim envia llavors una pàgina en format HTML sense que es pugui rebre el propi codi PHP que l'ha generada (tot i que amb enginyeria inversa es podria deduir).

Pòlissa de Privacitat (Privacy Policy): És un document legal que restringeix els usos i aprofitament de les dades dels clients d'un servei. La informació personal no pot ser una cosa que sigui utilitzada per identificar així com així a un individu incloent el seu nom, direcció, data de naixement, estat civil, dades financeres, historial mèdic,... En el món dels negocis sovint limita quines dades es recopilaran i es conservaran, o quines es mantindran confidencials i quines seran

199 Col·laboradors de Viquipèdia, "Model de preservació OAIS," Viquipèdia, l'Enciclopèdia Lliure, [//ca.wikipedia.org/w/index.php?title=Model_de_preservaci%C3%B3_OAIS&oldid=13346769](http://ca.wikipedia.org/w/index.php?title=Model_de_preservaci%C3%B3_OAIS&oldid=13346769) (descarregat 7 de maig de 2014)

200 L. McGuinness, Deborah, and Frank van Harmelen. "OWL Web Ontology Language Overview." W3C. <http://www.w3.org/TR/owl-features/> (consultat a 16 de maig de 2014)

201 Wikipedia contributors, "Web Ontology Language," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Web_Ontology_Language&oldid=592974321 (consultat a 16 de maig de 2014)

compartides o fins i tot venudes.²⁰²

Rastrejador web (*crawler*): Eina informàtica que facilita el procés de còpia d'una web. És a dir, permet rastrejar tots el sistema de carpetes i fitxers d'un servidor web i triar fins a quina profunditat volem fer-ne una còpia i desar-la en el nostre espai d'emmagatzematge triat. També anomenat aranya web o robot web, facilita la inspecció de pàgines web de forma metòdica i automatitzada, per tal de crear-ne la còpia i poder processar-les posteriorment amb un motor de recerca que indexi les pàgines proporcionant un sistema de recerca ràpid²⁰³.

Retuit (*Retweet*): Reenviament d'un tuit o missatge públic de Twitter. Com que Twitter no va implantar oficialment aquesta opció fins al cap d'uns anys, aquesta acció s'havia de fer manualment copiant literalment el missatge públic i afegint-hi un "RT" a l'inici abans de l'usuari indicat al que havíem copiat i del que estàvem reenviant el missatge públic.

Robots.txt: És un estàndard d'exclusió, que permet indicar regles i peticions per a que els cercadors web no indexin certs continguts²⁰⁴.

Ruby on Rails (*Ruby*): És un entorn de programació web optimitzat per facilitar la creació de pàgines web. El seu propòsit inicial fou el de facilitar i agilitzar la creació d'entorns web que treballin amb bases de dades²⁰⁵.

SDK (*Software Development Kit*): És un conjunt d'eines de desenvolupament de programari, que permet als programadors disposar d'indicacions sobre com crear certs paquets de programari per un servei en concret, una plataforma de maquinari o de programari. Pot disposar també d'una part de programació d'aplicacions concretes que s'anomenaria API²⁰⁶.

Snowflake (*Twitter*): Servei de la xarxa Twitter que genera numeració d'identificació a una escala el més alta possible i amb les garanties el més segures possibles²⁰⁷. Es va crear quan es va veure que MySQL no permetia generar aquesta numeració seqüencial a un nivell prou alt i per tant Twitter va haver-se de crear el seu propi sistema. Aquesta numeració havia d'encabir-se en 64 bits²⁰⁸.

202 Wikipedia contributors, "Privacy policy," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Privacy_policy&oldid=606483168 (consultat a 16 de maig de 2014)

203 Col·laboradors de Viquipèdia, "Aranya web," Viquipèdia, l'Enciclopèdia Lliure, http://ca.wikipedia.org/w/index.php?title=Aranya_web&oldid=11198847 (descarregat 9 de març de 2013)

204 "About /robots.txt." The Web Robots Pages. <http://www.robotstxt.org/robotstxt.html> (consultat a 16 de maig de 2014)

205 Col·laboradors de Viquipèdia, "Ruby on Rails," Viquipèdia, l'Enciclopèdia Lliure, http://ca.wikipedia.org/w/index.php?title=Ruby_on_Rails&oldid=11450328 (descarregat 16 de març de 2013)

206 Col·laboradors de Viquipèdia, "Equip de desenvolupament de programari," Viquipèdia, l'Enciclopèdia Lliure, http://ca.wikipedia.org/w/index.php?title=Equip_de_desenvolupament_de_programari&oldid=12809182 (descarregat 19 de gener de 2014)

207 "twitter/snowflake." GitHub. <https://github.com/twitter/snowflake/> (consultat a 16 de maig de 2014)

208 King, Ryan. "Announcing Snowflake" Twitter Blogs. <https://blog.twitter.com/2010/announcing-snowflake> (consultat a 16 de maig de 2014)

SQL (*Structured Query Language*): Llenguatge de programació dissenyat per gestionar dades en un sistema relacional de gestió de base de dades.

Stream: Flux de dades.

Tuit: Missatge públic amb un nombre de caràcters limitat, publicat instantàniament a Twitter. En català pot anomenar-se també piulada o piulet. Termcat prioritza piulada²⁰⁹, però hem preferit usar "tuit" en aquest treball perquè és la forma més breu i que ens és més comuna, i perquè va ser la que finalment va entrar com a traducció oficial a Twitter.

UTF-8: Codificació de caràcters que fa servir Unicode. Molt utilitzada en el sistema de transmissió d'informació a Internet. Té compatibilitat retroactiva amb ASCII. El seu ús s'ha ampliat a sistemes operatius, llenguatges de programació entre d'altres²¹⁰.

Variables (Informàtica): Conjunt d'opcions disponibles per a personalitzar les peticions, consultes i ordres que s'ideen i s'executen en un entorn informàtic.

Verified Account: Compte de Twitter que ha estat validat i verificat per Twitter. Permet establir que l'autor que escriu i publica missatges a través d'aquell compte ha aportat prou informació personal d'identificació per a que es pugui corroborar que està d'acord amb el que es publica en aquell perfil i a través d'aquell compte. Aquesta característica només està destinada a personatges que demostrin certa notorietat pública més enllà del nombre de seguidors que tinguin²¹¹.

Virtualització (Informàtica): Sistema que permet compartir una màquina física per executar diverses màquines virtuals²¹².

Web semàntica: Concepte que defineix l'objectiu de crear un mitjà universal d'intercanvi d'informació significativa a nivell humà i màquina alhora dels continguts web. El seu proclamador fou Tim Berners-Lee, fundador del consorci W3C i un dels pares d'Internet. Un dels exemples que formarien part d'aquest concepte és el llenguatge HTML5²¹³.

Wiki (Informàtica): Format de pàgina web que permet i possibilita l'edició completa dels continguts que apareixen a la web.

209 "La consulta del mes." TERMCAT.

http://www.termcat.cat/ca/Comentaris_Terminologies/Consultes_Terminologiques/80/ (consultat a 16 de maig de 2014)

210 Wikipedia contributors, "UTF-8," Wikipedia, The Free Encyclopedia, <http://en.wikipedia.org/w/index.php?title=UTF-8&oldid=607209025> (consultat a 16 de maig de 2014)

211 "FAQs about verified accounts." Twitter Help Center. <https://support.twitter.com/groups/31-twitter-basics/topics/111-features/articles/119135-about-verified-accounts> (consultat a 16 de maig de 2014)

212 Col·laboradors de Viquipèdia, "Virtualització," Viquipèdia, l'Enciclopèdia Lliure, [//ca.wikipedia.org/w/index.php?title=Virtualitzaci%C3%B3&oldid=13312600](http://ca.wikipedia.org/w/index.php?title=Virtualitzaci%C3%B3&oldid=13312600) (descarregat 1 de maig de 2014)

213 Col·laboradors de Viquipèdia, "Web semàntic," Viquipèdia, l'Enciclopèdia Lliure, [//ca.wikipedia.org/w/index.php?title=Web_sem%C3%A0ntic&oldid=13074123](http://ca.wikipedia.org/w/index.php?title=Web_sem%C3%A0ntic&oldid=13074123) (descarregat 17 de març de 2014)

WORM (*Write Once Read Many*): Que s'escriu una vegada però es pot llegir múltiples vegades. És a dir només escriptura i lectura, però no modificació del contingut. Descriu un suport d'emmagatzematge la informació del qual no pot ser modificada una vegada s'ha registrat per primera vegada, però en canvi pot ser llegida tantes vegades com es mantingui estable el suport.

XML (*eXtensible Markup Language*): Llenguatge que defineix unes normes per codificar formats que siguin llegibles tant per humans com per màquines. Està produït pel consorci W3C, entre d'altres especificacions i estàndards oberts.

ZIP: Format de fitxer comprimit, que permet comprimir sense pèrdua²¹⁴.

6 Bibliografia i fonts

6.1 Documentació de Twitter

- "API Terms of Service: Diff between Sept 5, 2012 and July 2, 2013." Twitter Developers. <https://dev.twitter.com/terms/api-terms/diff-20130702>
- "Developer Rules of the Road." Twitter Developers. <https://dev.twitter.com/terms/api-terms>
- "Documentation." Twitter Developers. <https://dev.twitter.com/docs>
- "Entities." Twitter Developers. <https://dev.twitter.com/docs/platform-objects/entities>
- "Tweets." Twitter Developers. <https://dev.twitter.com/docs/platform-objects/tweets>
- "Twitter Application Management." Twitter. <https://apps.twitter.com>
- "Twitter Developers." Twitter. <https://dev.twitter.com>
- "Twitter IDs, JSON and Snowflake Twitter Developers." <https://dev.twitter.com/docs/twitter-ids-json-and-snowflake>
- "Twitter welcomes Gnip to the flock | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2014/twitter-welcomes-gnip-to-the-flock>
- Costa, Jason. "New Withheld Content Fields in API Responses | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2012/new-withheld-content-fields-api-responses>
- GET search/tweets | Twitter Developers <https://dev.twitter.com/docs/api/1.1/get/search/tweets>
- Krikorian, Raffi. "Introducing Twitter Data Grants | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2014/introducing-twitter-data-grants>

²¹⁴ Wikipedia contributors, "Zip (file format)," Wikipedia, The Free Encyclopedia, [http://en.wikipedia.org/w/index.php?title=Zip_\(file_format\)&oldid=607878640](http://en.wikipedia.org/w/index.php?title=Zip_(file_format)&oldid=607878640) (consultat a 16 de maig de 2014)

- Roomann-Kurrik, Arne. "Introducing new metadata for Tweets | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2013/introducing-new-metadata-for-tweets>
- Sippey, Michael. "Changes coming in Version 1.1 of the Twitter API | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2012/changes-coming-to-twitter-api>
- Sippey, Michael. "Experience more with expanded Tweets | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2012/experience-more-with-expanded-tweets>
- Sippey, Michael. "Vine: A new way to share video | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2013/vine-a-new-way-to-share-video>
- Stone, Biz. "Tweet Preservation | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2010/tweet-preservation>
- Twitter / Twitter Privacy Policy <https://twitter.com/privacy>
- Twitter / Twitter Terms of Service <https://twitter.com/tos>
- Twitter / Twitter Terms of Service – Previous Terms of Service <https://twitter.com/tos/previous>
- Twitter Help Center | FAQs about Verified Accounts <https://support.twitter.com/groups/31-twitter-basics/topics/111-features/articles/119135-about-verified-accounts>
- Twitter Help Center | The Twitter Rules <https://support.twitter.com/articles/18311-the-twitter-rules>
- Vandor, Mollie. "Your Twitter archive | Twitter Blogs." Twitter Blogs. <https://blog.twitter.com/2012/your-twitter-archive>

6.2 NARA i Library of Congress

- "A REPORT ON FEDERAL WEB 2.0 USE AND RECORD VALUE ." National Archives Records Management Information Page. <http://www.archives.gov/records-mgmt/resources/web2.0-use.pdf>
- "Update: Twitter Archives, Library of Congress." Library of Congress. http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf
- "Update on the Twitter Archive at the Library of Congress." Library of Congress Blog. <http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/>
- Allen, Erin. "Update on the Twitter Archive at the Library of Congress." Library of Congress Blog. <http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/>

- Davis-Castro's, Carla. "Best Practices for Social Media Capture." National Archives Records Management Information Page. <http://www.archives.gov/records-mgmt/resources/socialmediacapture.pdf>
- Library of Congress. "PREMIS Data Dictionary for Preservation Metadata, Version 2.2 (July 2102) - premis-2-2.pdf." PREMIS: Preservation Metadata Maintenance Activity (Library of Congress). <http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>
- National Archives and Records Administration. "Bulletin 2014-02." National Archives and Records Administration. <http://www.archives.gov/records-mgmt/bulletins/2014/2014-02.html>
- Raymond, Matt. "How Tweet It Is!: Library Acquires Entire Twitter Archive." Library of Congress Blog. <http://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/>
- Raymond, Matt. "The Library and Twitter: An FAQ." Library of Congress Blog. <http://blogs.loc.gov/loc/2010/04/the-library-and-twitter-an-faq/>
- Social Media and Web 2.0 <http://www.archives.gov/social-media/>
- Society of American Archivists. "EAD Tag Library." EAD: Encoded Archival Description Version 2002 Official Site (EAD Official Site, Library of Congress). http://www2.archivists.org/sites/all/files/EAD2002TL_5-03-V2.pdf
- The Library and Twitter: An FAQ | Library of Congress Blog <http://blogs.loc.gov/loc/2010/04/the-library-and-twitter-an-faq/>

6.3 Altres articles

- "Dublin Core Metadata Element Set, Version 1.1." DCMI Home: Dublin Core© Metadata Initiative (DCMI). <http://dublincore.org/documents/dces/>
- "Guía de usos y estilo en las Redes Sociales del Gobierno Vasco." CAST.indd. http://www.irekia.euskadi.net/assets/a_documents/1218/Gui%CC%81a_de_usos_y_estilo_en_las_Red_Sociales_del_Gobierno_Vasco.pdf
- "Information on web archiving." The National Archives. <https://www.nationalarchives.gov.uk/webarchive/information.htm>
- "METS Example Documents." Metadata Encoding and Transmission Standard (METS) OfficialWeb Site. <http://www.loc.gov/standards/mets/mets-examples.html>
- "UK Government Web Archive: Twitter." The National Archives. <https://www.nationalarchives.gov.uk/webarchive/twitter.htm>
- AGELET, F. (1998). "Internet, un nou horitzó per als arxius. L'experiència de l'Arxiu

- Municipal de Barcelona". A: Lligall [Barcelona], núm. 13, p. 185-204. GRAELLS, J. (2011). "Les xarxes socials revolucionaran la professió d'arxiver". A: Lligall [Barcelona], núm. 32, p. 157 – 171
- Arda, Zeynep . " Yo soy mi imagen 2.0: Aproximación a formas contemporáneas de construir la identidad humana." Tesis Doctorals en Xarxa. <http://www.tdx.cat/handle/10803/81855>
 - Buckland, Michael (2013). «La qualitat de la informació al web». BiD: textos universitaris de biblioteconomia i documentació, desembre, núm. 31. <http://bid.ub.edu/31/buckland1.htm>
 - Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. Proceedings of the 20th international conference on World wide web - WWW '11: 675. <http://portal.acm.org/citation.cfm?doid=1963405.1963500>
 - Catalina, Oliver Vidal. "Ontologia de Twitter : El primer pas per inferir informació dels tweets." Repositori institucional - UOC. <http://openaccess.uoc.edu/webapps/o2/handle/10609/28324> [<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/28324/6/colivervTFC0114memoria.pdf>]
 - Domingo Arnal, Miguel Ángel, and Julià Minguillon. "Modelado, extracción y análisis de información del flujo de datos de Twitter." Repositori institucional - UOC. <http://openaccess.uoc.edu/webapps/o2/handle/10609/18961> [<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/18961/8/mdomingoarnTFM0113memoria.pdf>]
 - E. Locher, Anita, Aina Giones-Valls, Eli Ramírez, and Greda Ortiz. "Seguiu-nos: estratègies de captura de tuits en català." BiD: textos universitaris de biblioteconomia i documentació, desembre, núm. 29. <http://bid.ub.edu/29/locher1.htm>
 - Guia d'usos i estil a les xarxes socials de la Generalitat de Catalunya http://www.gencat.cat/xarxessocials/pdf/v1_guia_usos_xarxa_cat.pdf
 - Gupta, Aditi, and Ponnurangam Kumaraguru. "Credibility ranking of tweets during high impact events." Credibility ranking of tweets during high impact events. <http://dl.acm.org/citation.cfm?id=2185356>
 - Gupta, Manish, and Peixiang Zhao. 2012. Evaluating Event Credibility on Twitter. In Proceedings of SIAM International Conference on Data Mining, 153-164. http://www.cs.uiuc.edu/~hanj/pdf/sdm12_mgupta.pdf
 - Historicizing New Media: A Content Analysis of Twitter

<http://onlinelibrary.wiley.com/doi/10.1111/jcom.12030/abstract;jsessionid=1D0DEDEBB16234374A9486E98947D048.f01t03>

[<http://onlinelibrary.wiley.com/doi/10.1111/jcom.12030/pdf>]

- Jensen, Marion, Caswell Tom, Ball Justin, Duffin Joel, and Barton Rob. "TwHistory : Sharing History Using Twitter." Repositori institucional - UOC. <http://openaccess.uoc.edu/webapps/o2/handle/10609/4942> [http://openaccess.uoc.edu/webapps/o2/bitstream/10609/4942/6/Caswell_editat.pdf]
- Puigtobella, Bernat. "Evgeny Morozov: "Potser és l'hora de desprivatitzar Facebook i Google"." Núvol. <http://www.nuvol.com/noticies/evgeny-morozov-potser-es-lhora-de-nacionalitzar-facebook-i-google/>
- SEBASTIÀ, M., (1998). "Archiveros/infotecólogos: la Internet como motor del cambio de imagen de los archiveros". A: El archivo en el entorno cultural. Coslada: Comunidad de Madrid, p. 91-112. <http://anabad.org/images/publicaciones/anabad/xiiarchivosmunicipales.pdf>
- Soler Jiménez, Joan. "Diplomàtica.cat." Urkundendämmerung o el mito de la muerte documental. <http://diplomaticapuntcat.blogspot.com.es/2013/02/urkundendammerung-o-el-mito-de-la.html>
- TERRADELLAS PRAT, Enric. "L'ús de twitter en el món dels arxius." Lligall 33 i 34. Revista catalana d'arxivística 2011-2012. Miscel·lània. Associació d'Arxivers i Gestors de Documents de Catalunya

6.4 Llibres

- CRUZ MUNDET, José Ramón. Archivística. Gestión de documentos y administración de archivos. Alianza Editorial, 2012.
- SOLER, Joan. *Del bit al logos. Preservar documents electrònics a l'Administració local*. Col·lecció Estudis. Diputació de Barcelona, octubre de 2009, Barcelona.

7 Annexos

7.1 Manual per extreure dades en JSON amb Postman i l'API de Twitter

Com hem comentat varies vegades al treball, treballar amb l'API de Twitter és un procés que ha anat variant ja des de la versió 1.0 fins a l'actual 1.1. Als inicis tan sols necessitàvem conèixer la URL de la petició i les variables disponibles. Ara però se'ns requereix d'autenticacions i utilitzar un programa que envii ordres via HTTP als serveis web. Procés que s'ha fet més complex, alhora que més segur, i que a Twitter li permet un major control sobre qui accedeix al servei.

Per treballar amb la API de Twitter tenim diversos sistemes, biblioteques i llenguatges²¹⁵, però per aquest treball hem optat per la via més ràpida i senzilla, a més de ser la més accessible i assolible per a la majoria d'usuaris. Es tracta d'utilitzar l'extensió Postman per Google Chrome. Twitter recomanava usar un programa que es diu cURL²¹⁶, que vam provar de fer-lo funcionar, però en la seva versió per Windows ens donava problemes i en la seva versió per GNU/Linux ens anava millor, però el procés d'explicar uns instal·lació GNU/Linux més el de com treballar a nivell de línia d'ordres en un sistema UNIX ens suposava un procés massa complex com per recomanar-lo a nivell d'usuari comú. Per això quan vam conèixer l'existència de Postman vam decidir optar per recomanar-lo al ser-nos més assequible i no tant rebuscat.

Postman²¹⁷ és un programa que permet fer consultes via HTTP i provar les funcions de les API de serveis web. Funciona a través de menús i una interfície molt visual que no requereix gaires coneixements informàtics més enllà de saber què hem de posar a cada casella del programa i perquè funciona cada botó.

Primer però haurem de crear la nostra aplicació a Twitter per tal que ens faciliti l'accés a la API.

Ens dirigim a la web de desenvolupadors de Twitter (<https://dev.twitter.com>) i un cop allà entrem a dalt de tot a "Sign in" amb el nostre usuari de Twitter. Després quan ja haguem entrat ens dirigim a la nostra icona, cliquem i anem a "My applications", que ens portarà a Twitter Application Management (<https://apps.twitter.com>) i cliquem a "Create New App".

Twitter Apps



Il·lustració 4: Capçalera de la pàgina de Twitter Application Management.

Un cop a dins del formulari omplim els camps amb les dades que creguem convenients, i acceptem les condicions del servei.

²¹⁵ "Twitter Libraries." Twitter Developers. <https://dev.twitter.com/docs/twitter-libraries> (consultat a 17 de maig de 2014)

²¹⁶ "curl groks URLs." curl and libcurl. <http://curl.haxx.se> (consultat a 17 de maig de 2014)

²¹⁷ "Postman - A powerful HTTP client to test web services." Postman - A powerful HTTP client to test web services. <http://www.getpostman.com> (consultat a 17 de maig de 2014)

Create an application

Application details

Name *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.

(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Where should we return after successfully authenticating? [OAuth 1.0a](#) applications should explicitly specify their `oauth_callback` URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

Developer Rules of the Road

Last Update: July 2, 2013.

Twitter maintains an open platform that supports the millions of people around the world who are sharing and discovering what's happening now. We want to empower our ecosystem partners to build valuable businesses around the information flowing through Twitter. At the same time, we aim to strike a balance between encouraging interesting development and protecting both Twitter's and users' rights.

So, we've come up with a set of Developer Rules of the Road ("**Rules**") that describes the policies and philosophy around what type of innovation is permitted with the content and information shared on Twitter.

The Rules will evolve along with our ecosystem as developers continue to innovate and find new, creative ways to use the Twitter API, so please check back periodically to see the current version. Don't do anything prohibited by the Rules and talk to us if you think we should make a change or give you an exception.


If your application will eventually need more than 1 million user tokens, or you expect your [embedded Tweets](#) and [embedded timelines](#) to exceed 10 million daily impressions, you will need to talk to us directly about your access to the Twitter API as you may be subject to additional terms. Furthermore, applications that attempt to replicate Twitter's core user experience (as described in Section 1.5 below)

☒ Yes, I agree

Create your Twitter application

Un cop creada, si ens dirigim a Settings podrem completar les dades descriptives de l'aplicació, posant-hi el nom de l'organització que volem o afegir-hi una icona que haguem dissenyat.

Application icon

 **Change icon**

No s'ha seleccionat cap fitxer.

Maximum size of 700k. JPG, GIF, PNG.

Organization

Organization name

The organization or company behind this application, if any.

Organization website

The organization or company behind this application's web page, if any.

I quan ho tinguem tot omplert ens dirigim a Permissions i activem Read and Write.

Access

What type of access does your application need?

Read more about our [Application Permission Model](#).

☐ Read only

☒ Read and Write

☐ Read, Write and Access direct messages

Note:

Changes to the application permission model will only reflect in access tokens obtained after the permission model change is saved. You will need to re-negotiate existing access tokens to alter the permission level associated with each of your application's users.

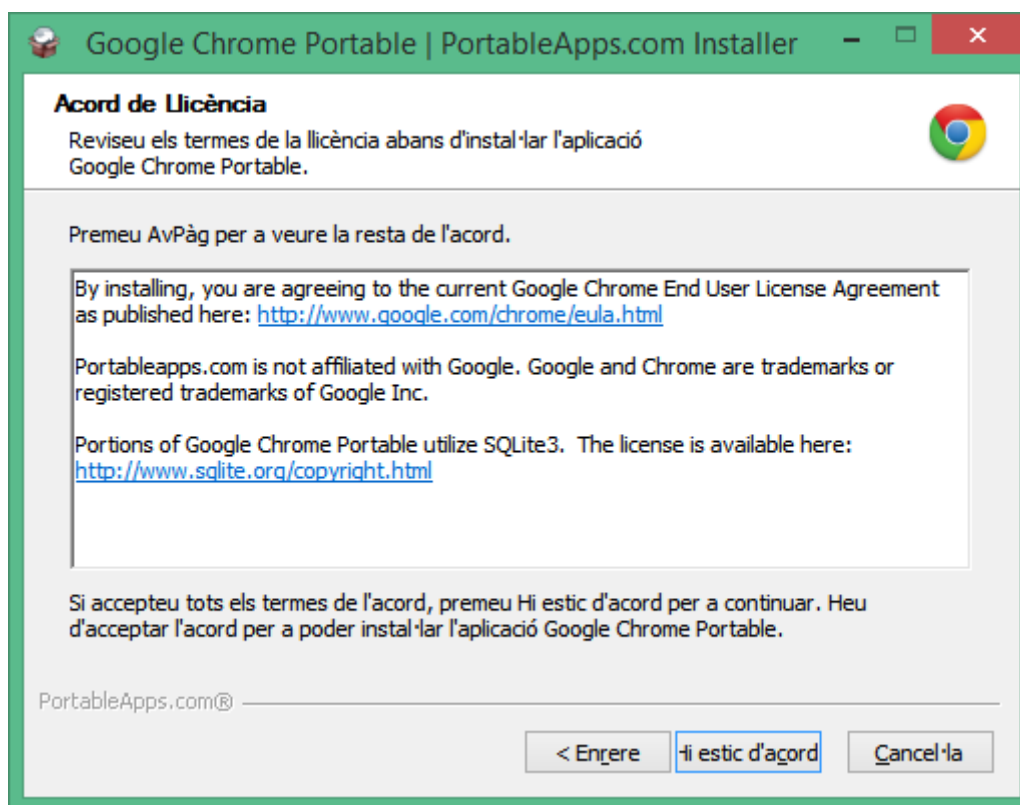
Ara que ja tenim l'aplicació creada i configurada, mantenim però la finestra oberta amb la pestanya del navegador oberta a *API Keys* que és la que ens servirà per configurar Postman. Haurem de tenir ben visible el que ens hi diu, sobretot la sèrie alfanumèrica de l'*API key* i l'*API secret*, com també *Access token* i *Access token secret*, ja que ens serviran per indicar a Postman el nostre accés a l'API.

El primer que necessitarem és tenir Google Chrome instal·lat al nostre sistema. Això és tan fàcil com anar a la web de Google Chrome²¹⁸ descarregar-nos-en l'instal·lador i instal·lar-lo. Una altra opció és anar a la web de PortableApps²¹⁹ i descarregar l'instal·lador d'una versió portable de Google Chrome la qual podrem copiar i moure allà on estiguem. Nosaltres hem optat per la versió

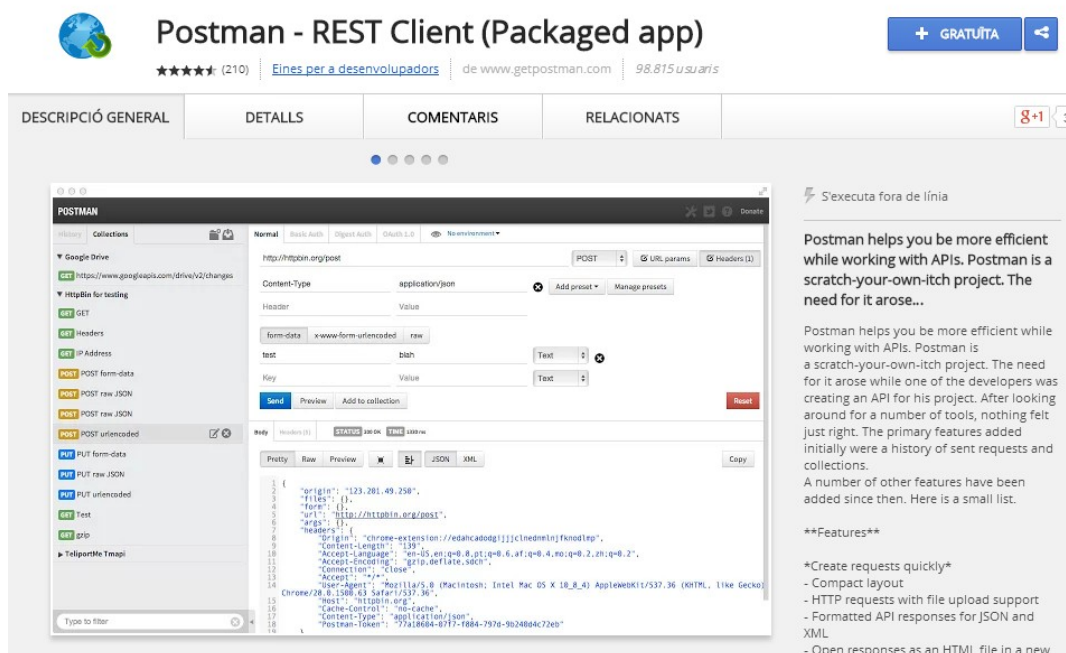
²¹⁸ "Un navegador web ràpid i gratuït." Navegador Chrome. <http://www.google.com/intl/ca/chrome/browser/> (consultat a 17 de maig de 2014)

²¹⁹ "Google Chrome Portable | PortableApps.com - Portable software for USB, portable and cloud drives." PortableAppscom News. http://portableapps.com/apps/internet/google_chrome_portable (consultat a 17 de maig de 2014)

portable de Google Chrome per no interferir amb l'ordinador en que treballem ja que és compartit i treballar amb la API de Twitter requereix l'ús d'informació sensible pel nostre usuari de Twitter.



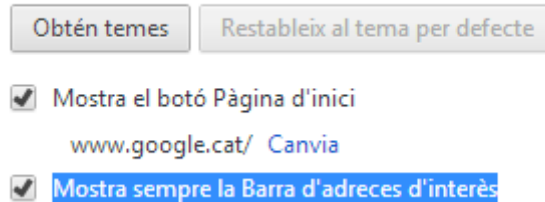
Un cop disposem de Google Chrome, ens dirigim a la web de Postman²²⁰ i en descarreguem l'extensió.



²²⁰ "Postman - REST Client." Chrome Web Store -. <https://chrome.google.com/webstore/detail/postman-rest-client/fdmmgilgnpjijdojopjoooidkmcomcm> (consultat a 15 de juny de 2014)

Un cop instal·lada aquesta extensió la trobarem a l'apartat Aplicacions de Google Chrome. I si volem tenir-ne un accés directe permanent podem dirigir-nos a la Configuració del navegador i activar l'opció "Mostra sempre la Barra d'adreces d'interès".

Aparença



Ara ens dirigim a Aplicacions des de la Barra d'adreces d'interès i un cop a dins cliquem sobre la icona de Postman.



Se'ns obrirà una finestra nova amb tot el menú de Postman. Ens dirigim a OAuth 1.0, i complimentem el formulari amb les dades que teníem a la pestanya oberta amb les *API keys* de Twitter. La correspondència de dades serà:

- Consumer Key - API key
- Consumer Secret - API secret
- Token - Access token
- Token Secret - Access token secret

No cal modificar cap més opció excepte marcar *Add params to header* a Postman i ja tindrem l'accés configurat. I cliquem a Refresh request.

Ens dirigim de nou a la pestanya Normal per indicar quina consulta volem fer. Obrim al nostre navegador la web de l'API de Twitter²²¹ i triem la consulta que desitgem, en aquest tutorial començarem amb la primera mateixa que ens surt a la llista, és a dir *GET statuses/mentions_timeline*. Aquesta funció ens retornarà les últimes respostes o mencions que haguem rebut al nostre usuari de Twitter.

²²¹ "REST API v1.1 Resources." Twitter Developers. <https://dev.twitter.com/docs/api/1.1> (consultat a 17 de maig de 2014)

GET statuses/mentions_timeline

View

What links here

Updated on Thu, 2013-06-20 14:39

API version 1.1

Returns the 20 most recent mentions (tweets containing a users's @screen_name) for the authenticating user.

The timeline returned is the equivalent of the one seen when you view [your mentions](#) on twitter.com.

This method can only return up to 800 tweets.

See [Working with Timelines](#) for instructions on traversing timelines.

Resource URL

https://api.twitter.com/1.1/statuses/mentions_timeline.json

Doncs bé la direcció *Resource URL* que ens indica és la que introduïrem a *Enter request URL here* a Postman. I com a variables hem optat per posar el límit count a 1 i hem desactivat que ens mostri informació el màxim de detallada dels usuaris.

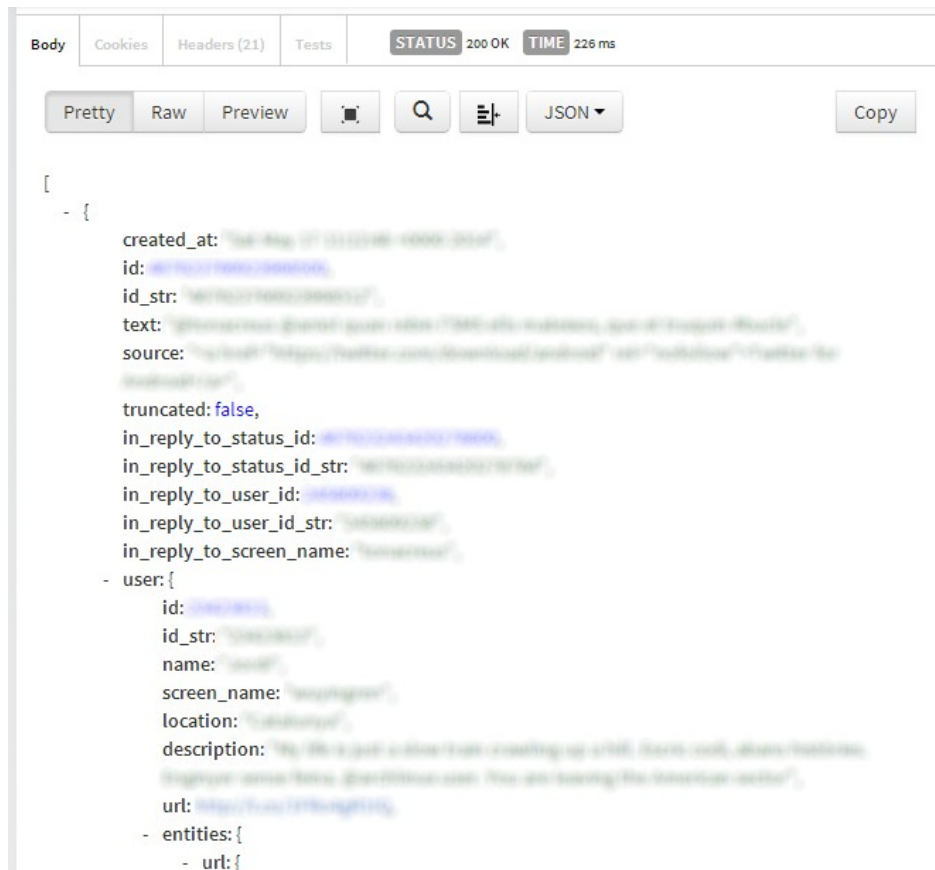
https://api.twitter.com/1.1/statuses/mentions		GET ▼	URL params	Headers (1)
count	1			✕
contributor_details	false			✕
URL Parameter Key	Value			
Authorization	OAuth realm="",oauth_consumer_ki			✕
Header	Value			

Add preset ▼
Manage presets

Send ▼
Preview
Tests
Add to collection

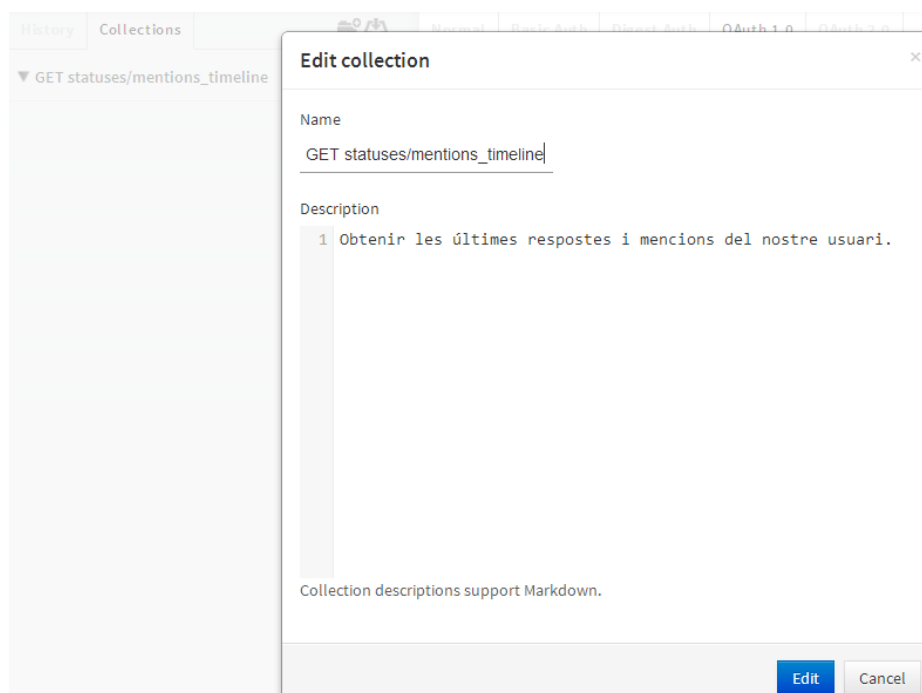
Reset

Seguidament al clicar a *Send* ens hauria de retornar el codi en JSON amb les dades que li hem demanat.



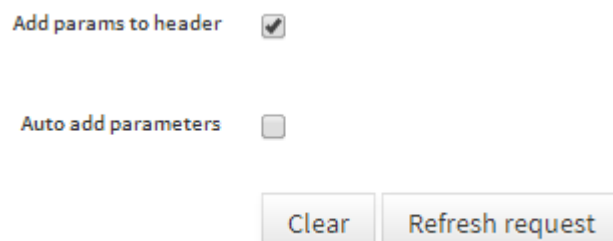
Tenim les opcions de tenir el codi en JSON, XML, HTML i Text, a més de visualitzar-lo amb ressaltat o en brut.

Amb això ja hem aconseguit fer una petició a l'API. Si ho desitgem, per finalitzar podem guardar la configuració d'aquesta petició a la barra esquerra a *Collections*.



Ara cada vegada que desitgem fer una consulta a l'API només cal que visitem la web de l'API de Twitter²²² i cerquem quina és la consulta que necessitem.

Hem de tenir present però que per cada nova consulta que fem ens pot sortir l'error 401. En aquest cas vol dir que ens hem oblidat de refrescar l'accés a Twitter, així que ens haurem de redirigir altra vegada a la pestanya de OAuth 1.0 i clicar a *Refresh request*.



Add params to header ☒

Auto add parameters ☐

Clear Refresh request

7.1.1 Les metadades d'un tuit

En aquest treball hem comentat que es podien extreure dades i metadades de l'API de Twitter d'un tuit concret. Doncs seguint el manual anterior on hem explicat com instal·lar i fer servir Postman, seguim amb un exemple de com fer una petició a l'API d'un tuit concret.

Primer de tot ens dirigim a l'apartat de l'API de Twitter *GET statuses/show/:id*²²³ on se'ns informa de la configuració i paràmetres per aquesta consulta d'un tuit.

Returns a single [Tweet](#), specified by the `id` parameter. The Tweet's author will also be embedded within the tweet.

See [GET statuses/lookup](#) for getting Tweets in bulk (up to 100 per call). See also [Embeddable Timelines](#), [Embeddable Tweets](#), and [GET statuses/oembed](#) for tools to render Tweets according to [Display Requirements](#).

Resource URL

<https://api.twitter.com/1.1/statuses/show.json>

Configurem Postman per a que ens ofereixi la consulta d'un tuit agafat a l'atzar²²⁴. Recordar que si ens surt l'error 401 que hem comentat al final de l'anterior apartat, hem d'anar a OAuth 1.0 i clicar a "Refresh request".

²²² "REST API v1.1 Resources." Twitter Developers. <https://dev.twitter.com/docs/api/1.1> (consultat a 17 de maig de 2014)

²²³ "GET statuses/show/:id." Twitter Developers. <https://dev.twitter.com/docs/api/1.1/get/statuses/show/%3Aid> (consultat a 17 de maig de 2014)

²²⁴ L'he triat amb certa ironia, però ha sigut a l'atzar dels que en el moment de redactar acabaven de ser publicats per Barack Obama (@barackobama).

https://api.twitter.com/1.1/statuses/show GET URL params Headers (1)

id	464853505052651520	✕
trim_user	true	✕
include_entities	false	✕
URL Parameter Key	Value	
Authorization	OAuth realm="",oauth_consumer_k	✕
Header	Value	

Add preset Manage presets

Send Preview Tests Add to collection Reset

Com es pot veure en la captura de pantalla hem activat que només rebem el número identificatiu de l'usuari i no pas totes les metadades que identificarien l'usuari que ha enviat el tuit. També hem desactivat que s'incloguin totes les entitats suplementàries del tuit tals com informació del perfil o informació de les direccions web, la ubicació de les imatges incrustades al tuit, etcètera. Tota aquesta informació la podem extreure de l'apartat *Parameters* de la pàgina concreta de l'API.

Returns a single [Tweet](#), specified by the id parameter. The Tweet's author will also be embedded within the tweet.

See [GET statuses/lookup](#) for getting Tweets in bulk (up to 100 per call). See also [Embeddable Timelines](#), [Embeddable Tweets](#), and [GET statuses/oembed](#) for tools to render Tweets according to [Display Requirements](#).

Resource URL

https://api.twitter.com/1.1/statuses/show.json

Llavors, quan cliquem a Send ens apareixerà tot el codi en JSON de manera resumida degut als paràmetres que hem establert en la consulta.

```

{
  created_at: "Fri May 09 19:44:46 +0000 2014",
  id: 464853505052651500,
  id_str: "464853505052651520",
  text: "Warm up. http://t.co/C5obglOr58",
  source: "web",
  truncated: false,
  in_reply_to_status_id: null,
  in_reply_to_status_id_str: null,
  in_reply_to_user_id: null,
  in_reply_to_user_id_str: null,
  in_reply_to_screen_name: null,
  user: {
    id: 813286,
    id_str: "813286"
  },
  geo: null,
  coordinates: null,
  place: null,
  contributors: null,
  retweet_count: 6412,
  favorite_count: 10038,
  favorited: false,
  retweeted: false,
  possibly_sensitive: false,
  lang: "en"
}

```

7.1.2 Les metadades d'un perfil

Seguim els mateixos passos que amb l'anterior apartat, anem a la pàgina concreta de l'API de Twitter *GET statuses/user_timeline*²²⁵ i agafem les dades que necessitem. En aquest cas ens sortiran els tuits publicats per un usuari concret, o si no en posem cap ens sortirien els nostres propis tuits.

screen_name	barackobama	✕
count	1	✕
trim_user	true	✕
exclude_replies	false	✕
contributor_details	false	✕
include_rts	true	✕
include_entities	false	✕
URL Parameter Key	Value	
Authorization	OAuth realm="",oauth_consumer_ki	✕
Header	Value	

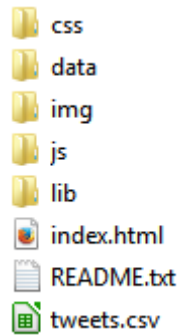
²²⁵ "GET statuses/user_timeline." Twitter Developers.

https://dev.twitter.com/docs/api/1.1/get/statuses/user_timeline (consultat a 17 de maig de 2014)

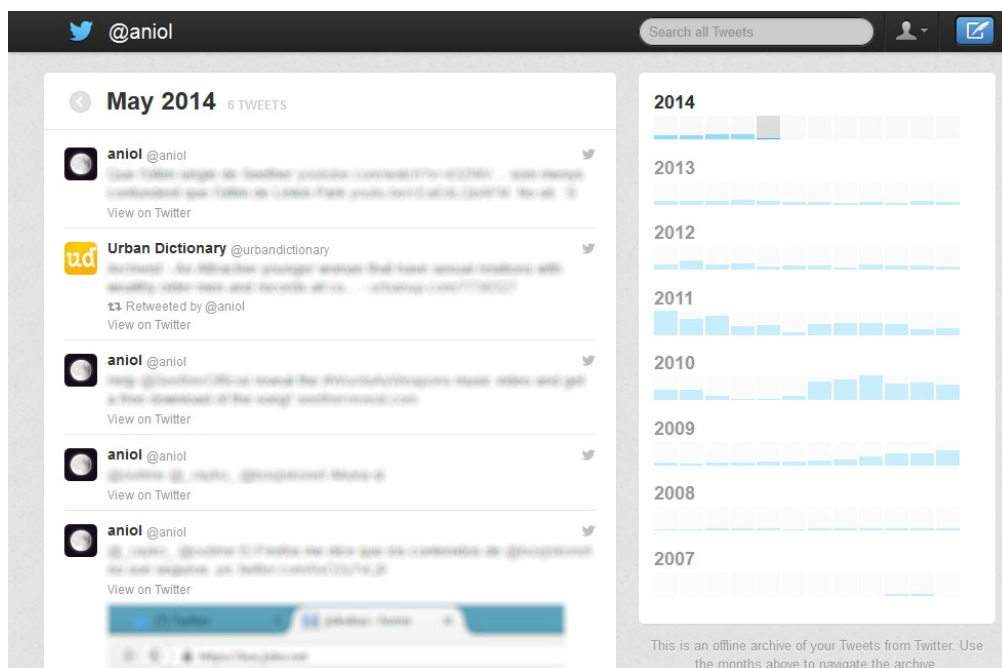
En aquest cas hem tornat a triar el perfil de Barack Obama (@barackobama), que podríem haver indicat pel valor numèric del seu usuari però per no complicar-nos el procés hem optat per posar directament el seu nom d'usuari. Hem indicat que volíem obtenir només un sol tuit (*count*), que volíem la informació de l'usuari resumida (*trim_user*), que no volíem les respostes que acompanyaven al tuit (*exclude_replies*), també hem desactivat la informació extra dels que han contribuït al tuit (*contributor_details*), hem activat però la informació sobre els retuits (*include_rts*), i finalment hem provat de desactivar-ne la informació sobre les entitats (*include_entities*) que malgrat no sortia aquesta opció a la pàgina d'informació de Twitter sembla que igualment funciona.

```
[
  - {
    created_at: "Sat May 17 16:59:30 +0000 2014",
    id: 467711016604667900,
    id_str: "467711016604667904",
    text: "Watch: In his weekly address, President Obama discusses actions to expand
    opportunity for more Americans. http://t.co/mbdR7XX27M",
    source: "web",
    truncated: false,
    in_reply_to_status_id: null,
    in_reply_to_status_id_str: null,
    in_reply_to_user_id: null,
    in_reply_to_user_id_str: null,
    in_reply_to_screen_name: null,
    - user: {
      id: 813286,
      id_str: "813286"
    },
    geo: null,
    coordinates: null,
    place: null,
    contributors: null,
    retweet_count: 111,
    favorite_count: 188,
    favorited: false,
    retweeted: false,
    possibly_sensitive: false,
    lang: "en"
  }
]
```

I aquí les tenim, les metadades en JSON del perfil de Barack Obama a Twitter segons les indicacions que li hem demanat a l'API de Twitter.



Si obrim el fitxer index.html, podrem visualitzar amb tota una interfície web creada per Twitter els nostres tuits de forma molt similar a com es visualitzen a la web de Twitter, oferint-nos a més a la columna dreta un histograma amb els anys i mesos en que hem publicat tuits.



El fitxer tweets.csv és el que anomenem sovint en aquest treball i que un cop obert amb un programa adequat (LibreOffice Calc per exemple) ens permetria tenir accés a tots els nostres tuits ordenats en una taula amb els camps més bàsics en columnes.

A1	tweet_id				
	A	B	C	D	E
1	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source
2	1400000000000000000			1400000000000000000	web
3	1400000000000000000			1400000000000000000	web
4	1400000000000000000			1400000000000000000	Tweet Button
5	1400000000000000000			1400000000000000000	web
6	1400000000000000000			1400000000000000000	web
7	1400000000000000000			1400000000000000000	twicca
8	1400000000000000000			1400000000000000000	web

Aquest fitxer CSV pot ser-nos molt útil si volem fer cerques i proves sobre un conjunt de tuits nostres. Sobretot m'ha sigut útil a mi per fer cerques de tuits publicats per mi mateix quan el sistema de cerca web de Twitter no m'acabava de trobar el que estava cercant.